DEVELOPING OLEAGINOUS YEAST AS MODEL ORGANISMS FOR LIPID PRODUCTION

BY

ANSHU DEEWAN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemical Engineering
with a concentration in Computational Science and Engineering
in the Graduate College of the
University of Illinois Urbana-Champaign, 2022

Urbana, Illinois

Doctoral Committee:

Professor Christopher V. Rao, Chair
Associate Professor Mary L. Kraft
Associate Professor Diwakar Shukla
Professor Yong-Su Jin

ABSTRACT

Over the last few decades, there has been a consistent research effort to reduce the dependence on fossil fuel-derived feedstocks. Plant biomass provides a sustainable alternative to crude-based feedstocks. Oleaginous yeasts (yeasts that can store 20% w/w lipids) are an emerging class of microorganisms, producing biofuels and bioproducts from renewable biomass. While yeasts like *Rhodosporiduim toruloides, Lipomyces starkeyi*, and *Yarrowia lipolytica* are ideal hosts for lipid production from biomass, we are limited in our understanding of the physiology of these yeasts. The work presented in this thesis aims at understanding the mechanisms involved in sugar utilization and lipid accumulation in oleaginous yeasts using various bioinformatics tools and mathematical modeling. This work improves our understanding of these yeasts and helps in the rational design of increased lipid-producing strains.

Chapter 1 provides a brief introduction to the fermentation of lignocellulosic hydrolysates. Chapter 2 reports the substrate uptake and utilization pathways in the oleaginous yeast *Rhodosporiduim toruloides*. In Chapter 3, we report the sugar utilization pathways in *Lipomyces starkeyi*. Using RNA sequencing and metabolomics, we identify the genes and pathways involved in the utilization of various sugars in both the yeasts. This work was done in collaboration with Dr. Sujit Jagtap and Dr. JingJing-Liu. Chapter 4 identifies the transporters involved in sugar uptakes in yeasts and compares them to known transporters in plants. We characterize the transporter *LST1_205437* from the yeast *Lipomyces starkeyi*, which is capable of transporting glucose and xylose simultaneously. This work was done in collaboration with Drs. Nurzhan Kuanyshev, Sujit Jagtap and JingJing-Liu. Chapter 5 discusses the insertion of a thioesterase gene in the yeast *Yarrowia lipolytica* for the production of medium chain alcohols. We present a bioinformatics pipeline to identify the gene's location, copy number variation, and

plasmid insertion. Chapter 6 summarizes this work and discusses the next steps toward developing these yeasts as model organisms.

Dedicated to my late uncle, Balveer Singh, for always cheering me on.

ACKNOWLEDGEMENTS

This thesis and my Ph.D. work would not have been a success without the support and encouragement of my advisor, Dr. Christopher V. Rao. Chris gave me the freedom to research at my own pace and was always available for advice and feedback when I needed it. Sharing his vast scientific knowledge, he has always encouraged me to strike a balance between learning and delivering in my research career, and I am thankful for his insights in my work. His mentorship was also invaluable to my career and life; I have grown both as a researcher and a person, in a big part because of his compassion. I have been fortunate to work with him for the last 6 years.

I want to acknowledge my fellow group members – past and present – who have provided a supportive work environment. Special thanks to Dr. Hanna Walukiewicz, Matt Plutz, and Dr. Sujit Jagtap for providing wet lab support and feedback through the early years of my Ph.D. I would also like to thank Dr. James Orr and Dr. Xiaoyi Wang for training me in the wet lab and answering my naïve questions. Thanks to Dr. JingJing Liu, Dr. Ashwini Bedekar, and Dr. Hyungi Koh for collaborating with me. I also thank William Woodruff, Danielle Meyer, and Ebin Joseph for making the office a fun workspace and offering support in challenging times. Lastly, my stay at UIUC would not be the same without Dr. Girija Bodhankar. We moved to the CU area from India the same semester and have been inseparable since. I have learned so much about research and life from you, and I would not be writing this document if you were not around cheering me on all these years. Thank you for having my back, Girija! I also appreciate all the pets of our research group, whose pics and stories were a source of joy throughout the years.

I was lucky to be a part of various collaborations during my work, and I will forever be grateful for the opportunities provided by CABBI and GSE teams. I want to acknowledge Dr. Yong-Su Jin, who has looked closely over my weekly research for the last few years. His

feedback vastly improved the scope of my research, and I find myself lucky to be under his tutelage. I am grateful for the new career opportunities that the Institute for Genomic Biology (IGB) provided. I would not be successful in my research without the support of the IGB Biotechnology center. Thanks to Dr. Chris Fields and his team for the workshops on bioinformatics, offering support for unconventional projects and advice for future career paths.

A part of my Ph.D. work was during the COVID-19 pandemic, and I appreciate the support I received from my research group and the team at CABBI to get through it. A special thanks to Elizabeth Murphy, who always looked out for us and our work-life balance alongside her administrative responsibilities. Thanks to Chris and Danielle for engaging in extended conversations about life outside the workplace and supporting my health and wellbeing during this time. Special thanks to Shekhar Mishra – we started as batchmates, became co-workers and close friends (in no particular order), and I have been fortunate to have the opportunity to engage with you and learn from you in research, life, and hobbies. Our honest conversations about mental health and well-being were a savior during this period.

I would also like to thank my parents for their love and compassion throughout my Ph.D. I have been lucky for their kindness, understanding, and support in my life. I would also like to acknowledge my friends from my old and new life – Sonali Ranga, Chandni Sikarwar, Dr. Emily Chen, Anjori Halder, Kayla Smith, Yuli Sung, and Nicole Kaczorek - for celebrating all the big and small wins and being there during the tough times.

Lastly, this work and my life would not be complete without Dr. Bijal Patel. Our journey has been the most exciting, from batchmates to friends to partners. I owe a ton of gratitude to his love and support. I am a better person for every day I have known him. I love you, Bijal.

TABLE OF CONTENTS

Chapter 1: Introduction

1.1 Storage metabolism

Microorganisms can utilize a variety of nutrients and thrive in a spectrum of environmental conditions. Carbohydrates are used as a primary energy source for cell maintenance and biomass production. However, depending on changes in the environment, carbon flux is diverted to different pathways, leading to different cellular processes and by-products (Fig. 1.1). A limitation of essential nutrients like nitrogen or phosphorus diverts the flux from cell growth to storage. The choice of storage metabolite depends on the organism, with glycogen, lipids, and poly-hydroxy butyrate being a few examples [1,2].

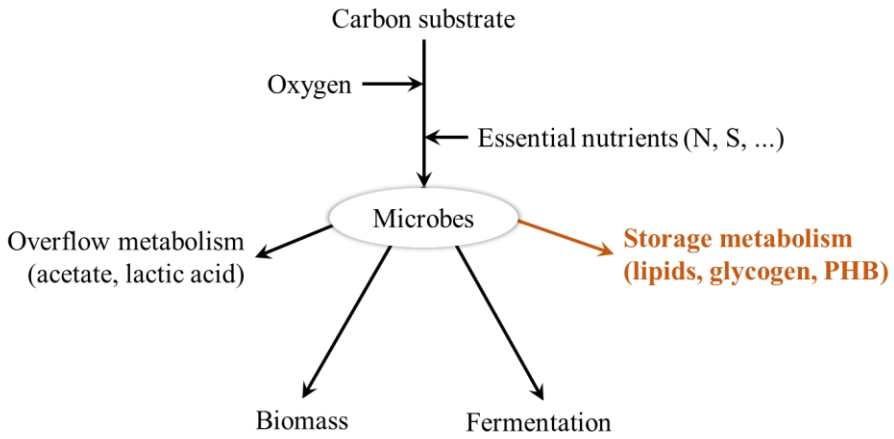

Figure 1.1. Pathways for carbon flux in microorganisms.

Storage metabolism is a phenomenon of interest for a variety of reasons. Storage metabolites are a microorganism's reserve to cope with starvation conditions present in the environment [3]. Lipids, produced in various eukaryotes, are involved in cellular homeostasis in oleaginous yeasts [4]. Lipids include a wide range of compounds, including fats, waxes, sterols, glycerolipids (monoglycerides, diglycerides, and triglycerides), phospholipids, and others. Bio-lipids have industrial applications in the areas of nutraceuticals, pharmaceuticals, fine chemicals, and fuels.

One of the most attractive utilizations of lipids is to generate biodiesel, which is considered a replacement for current diesel fuel [5].

1.2 Lignocellulosic hydrolysate as an alternative to petroleum derived feedstocks

A sustainable alternative to petroleum derivatives is plant biomass. Biomass from plants like miscanthus, sorghum, and switchgrass can be processed via pretreatment and utilized to produce biofuels, biochemicals, and biomaterials [6]. Lignocellulosic biomass is the non-edible part of the plant that primarily consists of cellulose, hemicellulose, and lignin. As lignocellulosic biomass does not compete for resources with food crops and arable land, they are a sustainable alternative for energy production. Using various thermal and biochemical conversion techniques, lignocellulosic biomass can be converted into hydrocarbons, alcohols, acids, and other organic compounds. Lignocellulosic biomass is also emerging as a valuable substrate for the production of value-added chemicals due to the chemistry of its major components, i.e., cellulose, hemicellulose, and lignin[7]. Various microorganisms have been used for the fermentation of lignocellulosic hydrolysate, for the production of fatty acids, triacylglycerides (TAG) and other lipids.

1.3 Lipid production by oleaginous yeasts

Storage metabolism can be induced in microorganisms by introducing nutrient stress like nitrogen or phosphorus starvation. Oleaginous yeasts are a class of fungi that have a natural capability to store high amounts of lipids (minimum 20% w/w), mostly in the form of triglycerides. *Yarrowia lipolytica*, which has been isolated from various oil rich media (like polluted sea water and dairy and meat products) is a well-studied oleaginous yeast. It can accumulate 30% of cell weight as lipids, and engineered strains have been reported to store 52% of cell weight as lipids[8]. *Lipomyces starkeyi* and *Rhodosporidium toruloides* are two other

oleaginous yeast of interest, which have been engineered to store 60% and 70% of cell weight as lipids, respectively[9]. Both these organisms can natively grow on a variety of sugars like glucose, xylose, cellobiose, and galactose, with are obtained from the hydrolysis of lignocellulosic biomass. While these two organisms are in focus for lipid production from renewable biomass, genetic tools are still being developed for their engineering. Their genomes have recently been sequenced, and the gene assembly is in progress[10,11].

Saccharomyces cerevisiae is one of the most intensively studied eukaryotic model organisms in molecular and cell biology. From an evolutionary perspective, *R. toruloides* is much further distant from *S. cerevisiae* as compared to *L. starkeyi* (Fig 1.2). Also, these oleaginous yeasts have metabolic pathways that are not present in *S. cerevisiae*. For instance, *S. cerevisiae* is a non-xylose utilizing organism, whereas *L. starkeyi* and *R. toruloides* can grow on xylose[12]. Another instance is the absence of ATP citrate lyase in *S. cerevisiae*, which converts cytosolic citrate to acetyl-CoA, which is a precursor to fatty acid biosynthesis[13]. Hence, using *S. cerevisiae* as a model species for genetic engineering of oleaginous yeasts is challenging. In order to harness their biosynthetic potential, it is important to understand the unique aspects of biosynthetic pathways in these oleaginous yeasts, specifically in the context of oleagincity and nutrient stress.

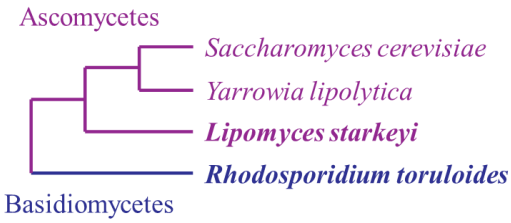

Figure 1.2. Phylogenetic tree of oleaginous yeasts of interest.

In recent work, transcriptomic and proteomic analysis of *R. toruloides* NP11 identified thousands of genes with altered transcript abundance under nitrogen and phosphorous limiting conditions[13,14]. In addition, genome sequences and draft gene models for *R. toruloides* IFO0880

and IFO0559 were generated[10,15]. However, work needs to be done to identify metabolic pathways and regulation during growth on multiple substrates. Gene regulation data would help understand substrate utilization mechanisms.

This work aims to understand molecular mechanisms involving lipid storage and nutrient utilization of oleaginous yeasts *R. toruloides* and *L. starkeyi* using transcriptomics and metabolomics. This involves studying the changes in metabolic pathways and substrate utilization of different carbon sources. This analysis further serves as a baseline for studying nutrient stress response (nitrogen or phosphorus starvation) in the cell. The second objective is to explain the effect of various nutrient starvations on storage metabolism and generates insights for improving lipid yields.

Chapter 2: Gene Regulation of Substrate Utilization in *Rhodosporidium toruloides*[1]

2.1 Introduction

Rhodosporidium toruloides is a red basidiomycete yeast that can produce a variety of biofuels and bioproducts from diverse carbon sources, including lignocellulosic hydrolysates and lignin-derived aromatics. [16–23]. It is an oleaginous yeast that accumulates triglycerides during growth under nitrogen, phosphorous, or sulfur limitation. [14,21,23,24]. In addition, *R. toruloides* can natively produce the sugar alcohols arabitol and galactitol from xylose and galactose, respectively.[25]. Recently, *R. toruloides* has been engineered for enhanced lipid production and production of fatty alcohols, terpenes, and methyl ketones [22,23,34,26–33]. Finally, *R. toruloides* can provide a natural source for carotenoids and industrially relevant enzymes such as L-phenylalanine ammonia lyase and D-amino acid oxidase [35–39].

Despite its ability to produce a wide range of value-added products, little is known about the physiology of this organism aside from a few key studies. In one seminal study, researchers measured changes in gene and protein expression in *R. toruloides* NP11 during growth on glucose in minimal medium under nitrogen limitation [13]. Based on these data, they were able to propose a model for lipid accumulation arising from nitrogen starvation. The same group also profiled changes in the lipid-droplet proteins in *R. toruloides* NP11 grown on glucose during nitrogen and phosphorous starvation [40]. In another study of note, phosphate limitation was found to increase the expression of genes involved in phosphate metabolism, RNA degradation, and lipid biosynthesis, while those involved in the tricarboxylic acid (TCA) cycle and ribosome biosynthesis were decreased [14]. In another seminal study, researchers constructed a barcoded

[1] This chapter is adapted from the following publication: Jagtap, S.S., **Deewan, A.**, Liu, JJ. *et al.* Integrating transcriptomic and metabolomic analysis of the oleaginous yeast *Rhodosporidium toruloides* IFO0880 during growth under different carbon sources. *Appl Microbiol Biotechnol* **105,** 7411–7425 (2021).

library of single-gene deletion mutants for *R. toruloides* IFO0880. Using this library, they were able to identify over a thousand essential genes and 150 genes affecting lipid production [41]. Recently, RNAseq analysis was performed on *R. toruloides* IFO0880 cultivated on synthetic defined medium on multiple carbon sources to identify the promoters that can constitutively express native genes at high and medium levels. In the process, the authors identified 12 mono-directional and 8 bi-directional promoters for *R. toruloides* [42]. Lastly, a recent study investigated the metabolic pathways and genes involved in carbon consumption in *R. toruloides* IFO0880 using multi-omics data. In addition, they have developed a genome scale metabolic network model by validating against growth phenotyping and gene fitness data [43].

In this work, we investigated the metabolism of glucose, xylose, acetate, and soybean oil by *R. toruloides* IFO0880. These substrates were chosen because they represent potential substrates for bioconversion by *R. toruloides*. To understand the metabolism of these substrates, we measured changes in gene expression using RNAseq and intracellular metabolites using GC-MS. We then mapped these data onto the central carbon metabolic pathways of *R. toruloides* in order to understand how they are regulated during growth on different carbon sources.

2.2 Materials and methods

Strains, media, and growth conditions

R. toruloides IFO0880, mating type A2, was obtained from the NITE Biological Resource Center in Japan (NBRC 0880). YPG medium (10 g/L yeast extract, 20 g/L peptone, and 20 g/L glucose) was used for growth of *R. toruloides*. A single colony from a YPG agar plate was inoculated into 2 mL of YPG liquid medium to obtain *R. toruloides* seed cultures. Seed cultures were then used to inoculate 25 mL of YPG, YPX medium (10 g/L yeast extract, 20 g/L peptone, and 20 g/L xylose), YPA medium (10 g/L yeast extract, 20 g/L peptone, and 20 g/L sodium

acetate, pH 7.0), YPS medium (10 g/L yeast extract, 20 g/L peptone, and 20 g/L soybean oil), and YP medium (10 g/L yeast extract and 20 g/L peptone) in a 125 mL baffled shake flask with a starting OD600 of 1. The optical density at 600 nm or OD600 was used to monitor the cell density in liquid cultures. OD600 of 1.0 corresponds to roughly 10^7 cells per mL. The cells were then grown at 30 °C and 250 rpm.

Sample extraction for RNA-seq and metabolomics

Seed cultures at exponential phase were collected and centrifuged at 6000 × *g* for 3 min at 4 °C. Supernatant was discarded and the pellets were resuspended in 1 mL of ddH_2O. Seed cultures then used to inoculate 25 mL of YPG, YPX, YPA, YPS, and YP medium in a 125 mL baffled shake flask with a starting OD600 of 1 and incubated at 30 °C and 250 rpm. Growth experiments are performed with three biological replicates. Samples from YPG, YPX, YPA, and YP media were collected after 24 h incubation, and sample from YPS media was collected after 16 h incubation.

Experimental procedure for RNA-Seq

The cell cultures containing a total OD of 30 were collected in centrifuge tubes and centrifuged at 6000 × g for 3 min at 4 °C. Supernatant was discarded and pellet was used for RNA extraction. Total RNA was extracted using the RNeasy mini kit (Qiagen, Hilden, Germany) as previously described, with a slight modification [44]. *R. toruloides* cell pellet was resuspended in 350 ul of Buffer RLT from the RNeasy mini kit (Qiagen, Hilden, Germany). Approximately 500 µl of acid-washed glass beads (acid washed, 425-600 µm; Sigma, St. Louis, MO, USA) were added and homogenized using a FastPrep-24 homogenizer (MP Biomedicals, Irvine, CA, USA), beaten at a speed of 5 m/s for 30 s six times with cooling on ice between beatings. The cell

lysates were purified according to the kit's protocol "purification of total RNA from yeast". Extracted RNA was then treated with Turbo RNase-free DNase kit (ThermoFisher, Waltham, MA, USA) according to the manual and purified again with the RNeasy mini kit protocol "RNA clean up". The stranded RNAseq libraries were prepared with Illumina's TruSeq Stranded mRNA Sample Prep kit. The libraries were quantitated by qPCR and sequenced on one lane for 101 cycles from one end of the fragments on a HiSeq 4000 (Illumina, San Diego, CA, USA). Fastq files with 100 bp reads were generated and demultiplexed with the bcl2fastq v2.17.1.14 Conversion Software (Illumina, San Diego, CA, USA). Raw sequencing files were uploaded to NCBI (NCBI Bioproject Accession: PRJNA660884, NCBI SRA Accession: SRR12567876 to SRR12567890).

RNA-Seq data analysis

To obtain gene expression profiles during growth of *R. toruloides* IFO0880 on different substrates, total RNA was extracted, and a mRNA focused library was sequenced. Adaptor sequences and low-quality reads were trimmed using Trimmomatic [45]. Trimmed reads were analyzed for quality scores using FastQC [46]. Reads were mapped to the *R. toruloides* IFO0880 v4.0 reference genome (NCBI Accession GCA_000988875.2) with STAR version 2.5.4a [41,47]. Between 95% and 98% of the reads were successfully mapped to the genome for each sample. Read counts were calculated using featureCounts from the Subread package, v1.5.2 [48]. Differential expression analysis was performed on the reads counts in R v4.0.5 using edgeR v3.32.1 and limma v3.46.0 [49,50]. Graphical representation of expression data was constructed using R packages: PCAtools v2.2.0, gplots v3.1.1, and Glimma v2.0.0 [51–53]. Before plotting heatmaps, the data was normalized row-wise (using the scale function in R), first by centering (subtracting the row mean from each value) and then scaling (dividing each data point

by row's standard deviation). Heatmaps were plotted using heatmap.2 function from gplots. Genome sequence, gene models and functional annotation of *R. toruloides* was downloaded from the DOE Joint Genome Institute's Mycocosm [41,54]. Data analysis scripts, along with the results, were uploaded at https://github.com/raogroupuiuc/rt88_growth.

Sample preparation for metabolome analysis

A fast filtration sampling method was used for intracellular metabolite analysis, as previously described, with a slight modification [55,56]. Briefly, 0.5 mL of cells were collected from all medias and vacuum-filtered using a vacuum manifold system (Vac-Man 96 Vacuum Manifold, Promega, Madison, WI, USA) assembled with a nylon membrane filter (0.45 µm pore size, 13 mm diameter, Whatman, Piscataway, NJ, USA) and a filter holder (Millipore, Billerica, MA, USA). The filtered cells were then washed with 2.5 mL of distilled water at room temperature. The entire process for fast filtration was completed within 1 min. The filter membrane containing the washed cells was quickly mixed with 1 mL of the pre-chilled acetonitrile/water mixture (1:1, v/v) and 100 µL of glass beads. The extraction mixture was vortexed for 3 min to disrupt the cell membrane and to extract intracellular metabolites. The extraction mixture was then centrifuged at 16,000 × *g* for 3 min at 4 °C, and 0.8 mL of the supernatant containing the intracellular metabolites was dried using a speed vacuum concentrator for 6 h. The samples were prepared for GC/MS analysis as mentioned in the analytical methods section.

Metabolome data analysis

The raw data obtained from the GC/MS analysis were processed using an automated mass spectral deconvolution and identification system (AMDIS) software for peak detection and deconvolution of mass spectra [57]. The processed data were uploaded to SpectConnect

(http://spectconnect.mit.edu) for peak alignment and generating the data matrix with the Golm Metabolome Database (GMD) mass spectral reference library [58,59]. The normalized abundance values for each metabolite were obtained by dividing the peak intensities with dry cell weight. For statistical analysis, such as PCA analysis and hierarchical cluster analysis representing as a heat map, Statistica (version 7.1; StatSoft, Tulsa, OK, USA), MetaboAnalyst and MultiExperiment Viewer software were used, respectively [60,61]. GraphPad Prism 6 (GraphPad, San Diego, CA, USA) was used for plotting intensity graphs.

Analytical methods

Prior to analysis, culture samples were centrifuged, and the supernatant was passed through a 0.22 μm polyethersulfone syringe filter. Glucose, xylose, and acetate concentrations were measured using a Shimadzu high performance liquid chromatography system (Shimadzu, Kyoto, Japan) equipped with a cation H micro-guard cartridge (Bio-Rad Laboratories Hercules, CA, USA) and Aminex HPX-87H carbohydrate analysis column (Bio-Rad Laboratories, Hercules, CA, USA) operated at 65 °C. Eluent was 5 mM H_2SO_4 at a constant flow rate of 0.6 mL/min. Compounds were monitored using a RID-10A refractive index detector (Shimadzu, Kyoto, Japan) and were quantified using calibration curves built using authentic standards.

The metabolome samples were derivatized by methoxyamination and trimethylsilylation as described previously [55,56]. For methoxyamination, 10 μL of 40 mg/mL methoxyamine chloride in pyridine (Sigma-Aldrich, St. Louis, MO, USA) was added to the samples and incubated for 90 min at 30 °C and 200 rpm. The samples were then trimethylsilylated by adding 45 μL of *N*-methyl-*N*-trimethylsilyltrifluoroacetamide (Sigma-Aldrich, St. Louis, MO, USA) for 30 min at 37 °C and 200 rpm. For GC/MS, the derivatized metabolite samples were applied to an Agilent 7890A GC/5975C MSD system (Agilent Technologies, Wilmington, DE, USA) equipped with a

RTX-5Sil MS capillary column (30 m × 0.25 mm, 0.25 μm film thickness; Restek, Bellefonte, PA, USA) and an additional 10 m long integrated guard column. One microliter of the derivatized sample was injected into the GC inlet in splitless mode. The oven temperature was initially set at 150° C for 1 min, after which the temperature was increased to 330°C at 20 °C/min, where it was held for 5 min. The mass spectra were recorded in a scan range 85–500 *m/z* at an electron impact of 70 eV, and the temperatures of the ion source and transfer line were 230 °C and 280 °C, respectively.

2.3 Results

Growth profiles in different carbon source

R. toruloides was grown on YP medium alone or on YP medium containing one of the following substrates: glucose (YPG), xylose (YPX), acetate (YPA), or soybean oil (YPS). YP was chosen as the base medium because it can support robust, vegetative growth. We note that *R. toruloides* can grow on YP medium alone. In addition, since this medium contains sufficient nutrients, the cells do not enter the lipogenic phase where lipids accumulate within the cell. The rationale for choosing this medium was that we were focused solely on exploring the metabolic pathways for utilizing these substrates rather than on identifying the pathways involved in lipid production. As shown in Fig. 2.1, *R. toruloides* achieved highest cell densities on glucose, followed by acetate, xylose, and soybean oil. Growth on soybean oil was similar to YP medium alone. In addition, *R. toruloides* was able to completely utilize 20 g/L glucose, xylose, and acetate after 48 h of growth. Consistent with the growth data, glucose was utilized at a rate faster than either xylose or acetate. Unfortunately, we were unable to accurately measure the utilization of soybean oil.

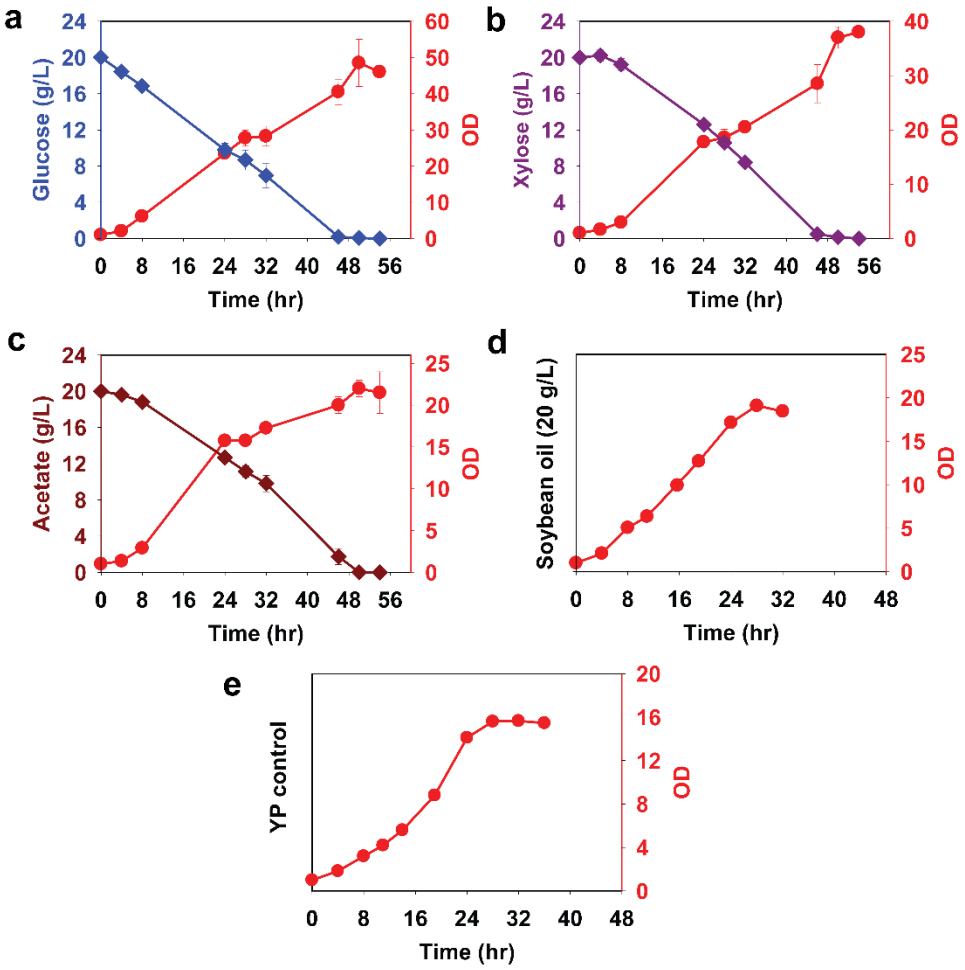

Figure 2.1. Growth profiles of *R. toruloides* IFO0880. Growth on different carbon sources at 20 g/L in rich medium: glucose **(a)**, xylose **(b)**, acetate **(c)**, soybean oil **(d)** and yeast peptone **(e)**. Solid squares (■) are used to denote sugar concentrations. Solid circles (●) are used to denote the optical density (OD600). For each plot (a-c), sugar concentration is plotted on the left y-axis, and OD600 values are plotted on the right y-axis.

Identification of genes associated with substrate utilization

We used RNA-Seq to measure differences in gene expression during growth on the different substrates. RNA was isolated for three biological replicates during the exponential growth phase (24 h for YPG, YPX, YPA, and YP medium; and 16 h for YPS medium). A total of 229 million raw reads were obtained from fifteen samples. Approximately 88% of the reads mapped to a unique location on the *R. toruloides* IFO0880 genome. Principal component analysis demonstrated that these growth changes led to clearly distinct expression patterns (Fig. 2.2).

Only during growth on YPS medium did we not observe a large change from growth on YP medium alone. This may reflect that growth on both was similar (Fig. 2.1).

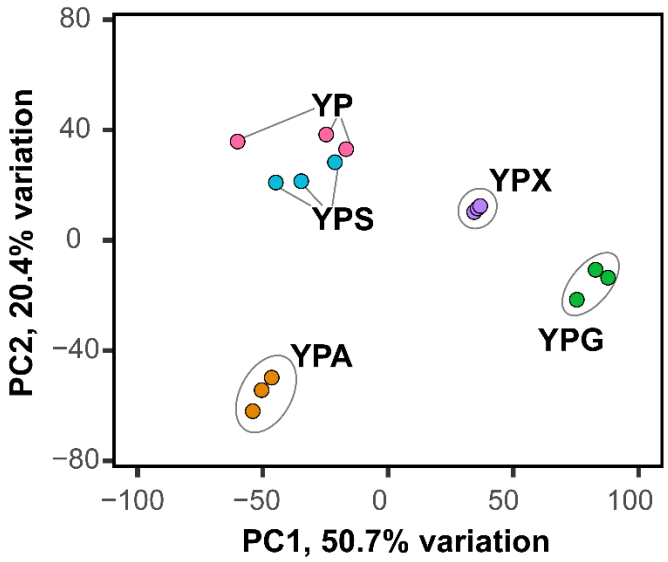

Figure 2.2. Principal component analysis plot generated from gene expression profiles of *R. toruloides* IFO0880 grown on glucose (YPG), xylose (YPX), acetate (YPA), soybean oil (YPS) and yeast peptone (YP). RNAseq data was collected in triplicate for each condition.

We next identified genes with altered expression during utilization of the different substrates. We compared growth on YPG with growth on YP to identify the pathways involved in uptake of glucose. We chose the same timepoint (24 h) for RNA extraction for both YPG and YP, with the intention of emphasizing the gene expression resulting from glucose uptake. For analyzing the gene expression patterns during growth on xylose, acetate, and soybean oil, growth on glucose was chosen as the control. YPG, YPX and YPA were sampled at 24 h, which was mid-exponential and close to 50% substrate utilization. However, YPS was sampled at 16 h, at mid-exponential growth (Fig. 2.1d).

R. toruloides IFO0880 has 8490 predicted genes. Among these genes, 3582 were upregulated and 3627 were downregulated during growth on the four substrates as compared to YP medium

alone. We also compared differential expression during growth on glucose, xylose, acetate, and soybean oil. 1488 genes have significantly higher expression, and another 1419 genes have significantly lower expression during growth on glucose as compared to YP. The expression of 882, 1297, and 1296 genes was significantly higher during growth on xylose, acetate, and soybean oil, respectively, as compared to glucose. Likewise, the expression of 537, 1693, and 1392 genes was significantly lower during growth on xylose, acetate, and soybean oil as compared to YPG. For comparative gene expression, fold change > 2 and adjusted p value < 0.05 was considered significant.

Changes in intracellular metabolites during growth on different substrates

We used gas chromatography-mass spectrometry (GC-MS) to measure changes in the concentrations of 55 intracellular metabolites as described previously. Among the 55 metabolites analyzed, significant differences in the concentrations of 44 metabolites were observed during growth on glucose, xylose, acetate, or soybean oil as compared to YP (Fig. 2.3). Significance was evaluated using two sample *t*-tests, and the cutoffs for significance were relative metabolite concentration > 2 and adjusted p value < 0.05. 16 metabolites had higher concentration and another 10 metabolites had lower concentrations during growth on glucose, as compared to YP. The concentrations of 7, 9, and 7 metabolites were higher during growth on xylose, acetate, and soybean oil, respectively, as compared to glucose. Likewise, the concentrations of 17, 11 and 16 metabolites were lower during growth on xylose, acetate, and soybean oil, respectively, as compared to glucose.

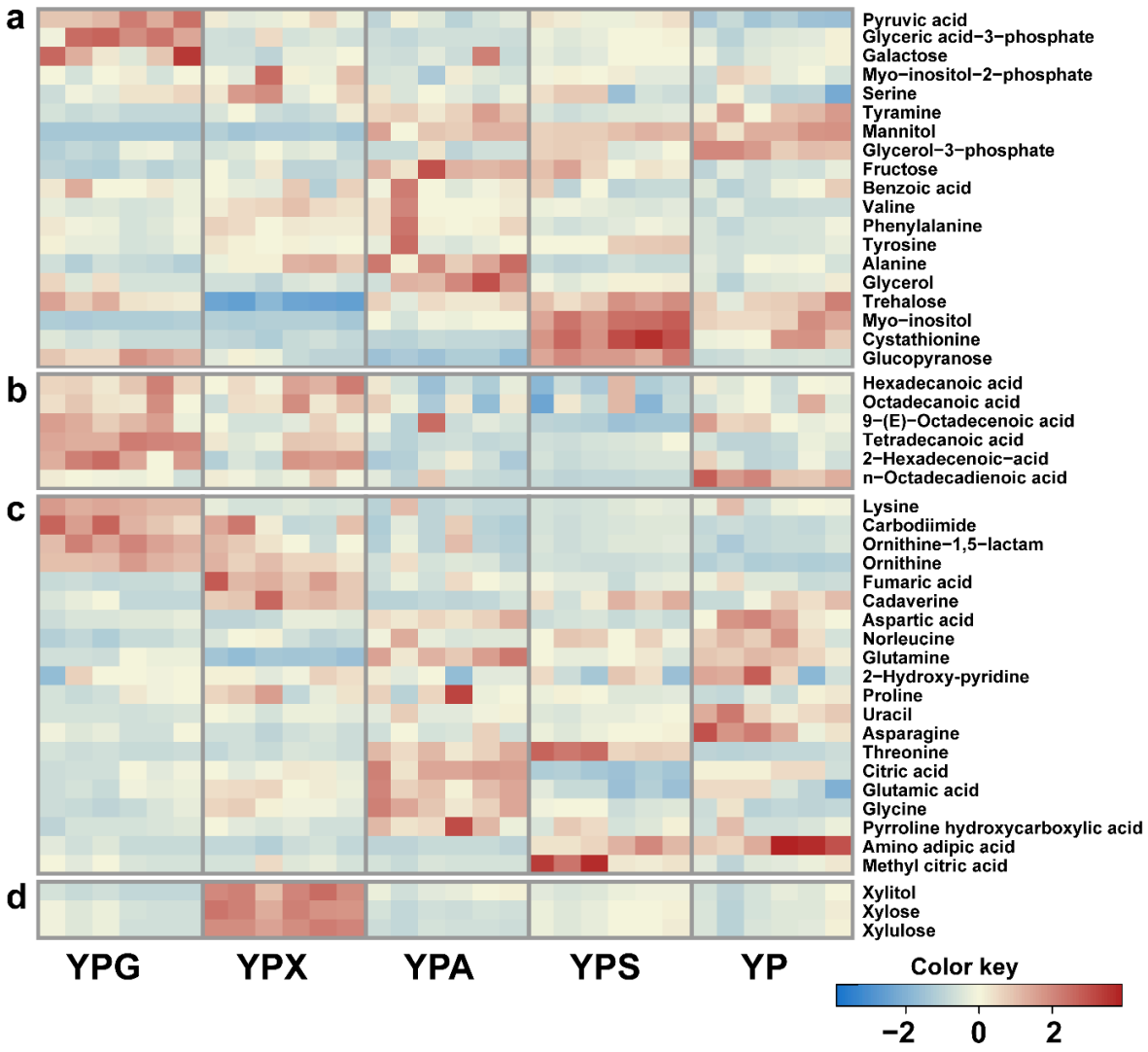

Figure 2.3. Heat map of intracellular metabolites from glycolysis **(a)**, fatty acid biosynthesis **(b)**, TCA cycle **(c)**, and xylose utilization **(d)** of *R. toruloides* IFO0880 grown on glucose, xylose, acetate, soybean oil and yeast peptone. All experiments were performed with six replicates. Color key represents the z-score for each metabolite (normalized for all growth conditions).

Growth on glucose

We first explored glucose metabolism in *R. toruloides* by comparing gene expression on YP medium containing glucose (YPG) versus YP medium alone. We chose the same timepoint (24 h) for analysis in order to isolate changes in gene expression resulting from glucose consumption. Overall, 1488 genes were significantly upregulated, and another 1419 genes were

significantly downregulated during growth on YPG versus YP alone (fold change > 2 and $p < 0.05$ was considered significant).

We focused on the specific genes involved in central carbon metabolism (Fig. 2.4 & 2.5). With regards to glycolysis, namely the Embden–Meyerhof–Parnas pathway, we found that the genes involved in the lower half of the pathway were upregulated at least 2-fold during growth on YPG versus YP. However, the expression of hexokinase (*HXK2*) and phosphofructokinase (*PFK*) was reduced during growth on YPG. We did not observe any significant changes in the expression of the genes involved in the pentose phosphate pathway except for 6-phosphogluconolactonase (*SOL3*), whose expression was increased during growth on YPG. In addition, the expression of the the alpha and beta subunits of pyruvate dehydrogenase (*PDA1* and *PDB1*) was increased during growth on YPG. Invertase expression (*SUC2*) was significantly reduced (32-fold) during growth on YPG, which is consistent with results from *S. cerevisiae* [62,63].

Significant changes were also observed in the expression of the genes involved in the TCA cycle. The expression of citrate synthase (*CIT1*), malate dehydrogenase (*MDH1*), and fumarate hydratase (*FUM1*) was increased during growth on YPG. However, the expression of succinate dehydrogenase (membrane anchor; *SDH4*) and α-ketoglutarate dehydrogenase (*KGD1*) was reduced. In addition, the expression of the cytosolic isocitrate lyase (*ICL1*) and malate synthase (*MLS1*) was reduced.

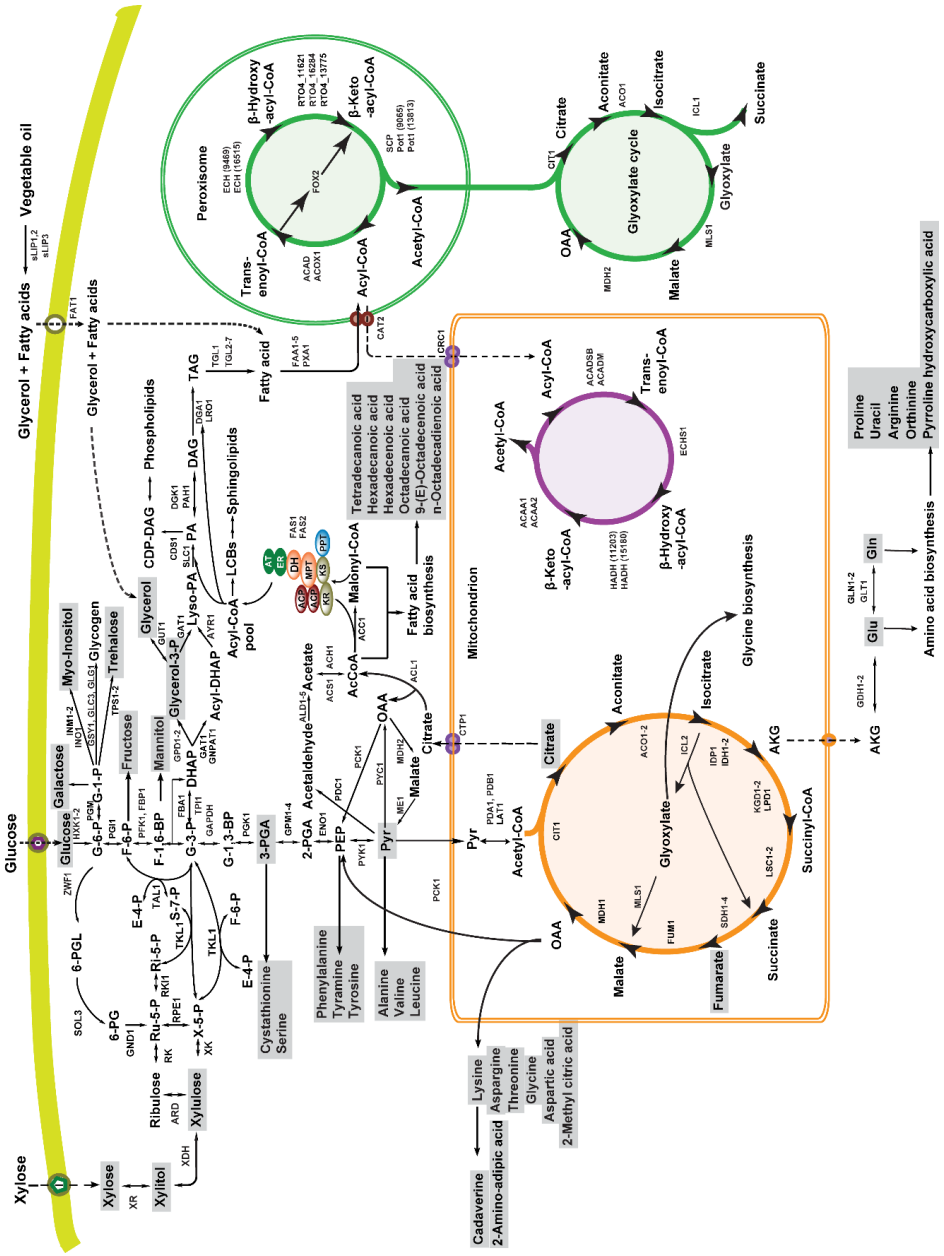

Figure 2.4. Central metabolic pathway in *R. toruloides* IFO0880. Intracellular metabolites reported in this study are highlighted in grey.

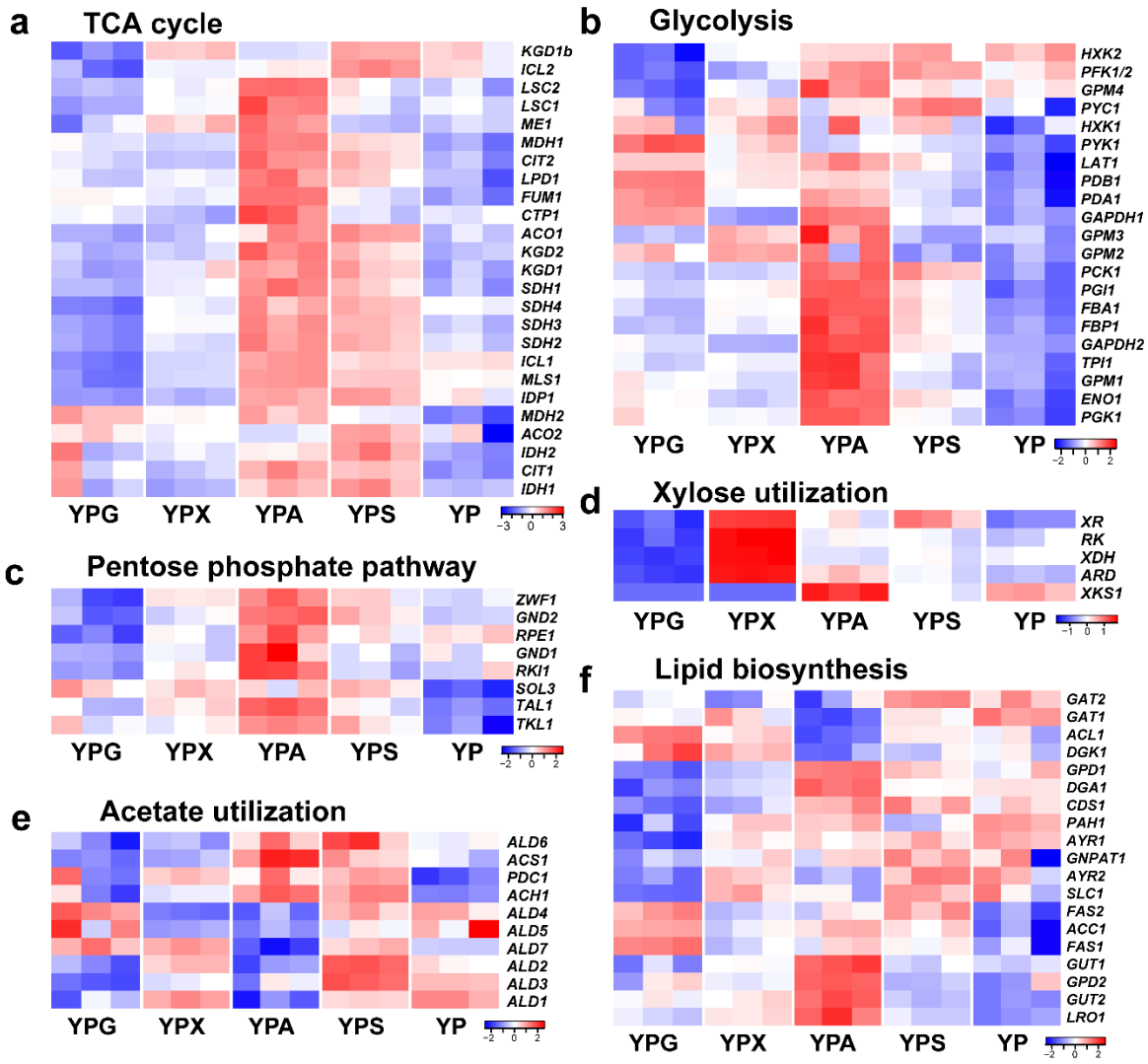

Figure 2.5. Transcriptional changes in pathways associated with TCA cycle **(a)**, glycolysis **(b)**, pentose phosphate pathway **(c)**, xylose utilization **(d)**, acetate utilization **(e)** and lipid biosynthesis **(f)** in *R. toruloides* IFO0880 grown on glucose (YPG), xylose (YPX), acetate (YPA), soybean oil (YPS) and yeast peptone (YP). RNAseq data was collected in triplicate for each condition. Color key represents the z-score for each gene (normalized for all growth conditions).

Aside from glucose metabolism, expression of the genes involved in lipid catabolism was significantly reduced. Expression of the genes involved in lipid biosynthesis was increased. These included ATP citrate lyase (*ACL1*), which provides cytosolic acetyl-CoA for lipid biosynthesis, and acetyl-CoA carboxylase, which generates malonyl-CoA. In addition, the expression of glutamate dehydrogenase (*GDH1*) was greatly increased (115-fold). However, the

expression of NAD-dependent glutamate dehydrogenase *GDH2* was reduced. The expression of general amino-acid GAP1 permeases (protein ID: 14229 and 15074) was also reduced. However, the expression of DIP5 dicarboxylic amino-acid transporters (protein ID: 8962, 9319, and 9322) and K16261 YAT amino-acid transporter (protein ID: 12758) was increased. We also observed the increased expression of several putative sugar transporters and the decreased expression of a few other putative sugar transporters. Whether these transporters are specific for glucose or other sugars is not known.

With respect to intracellular metabolites (Fig. 2.3), the concentrations of glucose, 3-phosphoglycerate, and pyruvate were increased during growth on YPG versus YP, which is consistent with their role in glycolysis. In addition, galactose concentrations were also increased. We also observed increased concentration of ornithine-1,5-lactam (more commonly known as 3-aminopiperidin-2-one), a delta lactam of ornithine. Consistent with the increase in expression of the genes involved in lipid synthesis, we observed increased concentrations of 2-hexadecenoic-acid, tetradecanoic acid, *n*-heptadecan-1-ol, and *n*-docosan-1-ol. The intracellular concentration of several amino acids – phenylalanine, lysine, threonine, and tyrosine –increased during growth on YPG. Interestingly, the concentration of asparagine decreased along with 2-amino-adipic acid and cystathionine, which are intermediates in lysine and cysteine biosynthesis, respectively. In addition, the concentration of tyramine, a byproduct of tyrosine, also decreased. Finally, we observed decreased concentration of the sugar alcohols: threitol, xylitol, and mannitol.

Growth on xylose

We next explored xylose metabolism in *R. toruloides* by comparing gene expression during growth in YP medium containing xylose (YPX) versus YP medium containing glucose (YPG). Overall, 882 genes were significantly upregulated and 537 genes were significantly

downregulated during growth on YPX versus YPG. The gene expression profile on xylose was the closest to growth on glucose, in comparison to the other substrates we tested.

Consistent with a mechanism where growth on xylose induces expression of the xylose metabolic genes, expression of xylose reductase *(XR*) and xylitol dehydrogenase (*XDH*) was increased during growth on YPX versus YPG. However, expression of xylulokinase (*XKS1)* was identical during growth on YPX and YPG. Further exploration of the data revealed that *XKS1* was not expressed during growth on either YPG or YPX. These results are consistent with recent observations [43]. This would suggest an alternate mechanism for xylose utilization. One possibility is the conversion of xylulose to arabitol, catalyzed by arabitol dehydrogenase (*ARD1*). Indeed, significant arabitol production was previously observed during growth on xylose in rich medium [64]. In support of this mechanism, expression of arabitol dehydrogenase (*ARD1*) was significantly higher (42-fold) during growth on YPX as compared to YPG. This enzyme converts xylulose to arabitol using NADH as the cofactor [65,66]. This enzyme can also convert arabitol to ribulose using NAD as the cofactor [66]. Ribulose is then phosphorylated by ribulose kinase (RK), forming ribulose-5-P. This metabolite is a substrate for the non-oxidative pentose phosphate pathway. Consistent with this mechanism, expression of the ribulose kinase gene (*RK*) was increased during growth on YPX as compared to YPG. Why *R. toruloides* employs this more circuitous route for xylose metabolism is not known.

Aside from enzymes directly involved in xylose utilization, we observed a few significant changes in the expression of genes involved in central carbon metabolism. Oxoglutarate dehydrogenase (*KGD1b*) was 27-fold upregulated. Pyruvate carboxylase (*PYC1)*, succinate dehydrogenase (*SDH4*), malic enzyme (*ME1*), isocitrate lyase (*ICL1*, *ICL2*), malate synthase (*MLS1*), and glucose-6-phosphate dehydrogenase (*ZWF1*) were all upregulated at least 2-fold

during growth on YPX as compared to YPG. Acetyl-CoA carboxylase (*ACC1*), ATP: citrate lyase (*ACL1*), and fatty-acyl-CoA synthase (*FAS1* and *FAS2*) were downregulated 2 to 3-fold.

From the metabolomics analysis , large increases in the concentrations of xylose (77-fold), xylitol (122-fold), and xylulose (3-fold) were observed, which is consistent with the XR/XDH pathway for xylose utilization. In addition, we also observed an increase in the concentration of xylonic acid (3-fold). This compound is formed from xylose by glucose oxidase [67]. However, we could not identify any gene associated with glucose oxidase activity. The concentrations of many intracellular acids decreased during growth on YPX, which is consistent with the downregulation of genes in the fatty acid biosynthesis pathway. In addition, we observed decreased concentrations of the storage carbohydrate trehalose, which is linked to glucose metabolism, as well as galactose and glycerol.

Growth on acetate

We next explored acetate metabolism in *R. toruloides* by comparing gene expression during growth on YP medium containing acetate (YPA) versus YPG medium. Overall, 1297 genes were significantly upregulated, and 1693 genes were significantly downregulated during growth on YPA versus YPG.

Acetate enters the central metabolism through the action of acetyl-CoA synthetase (ACS1) (Fig. 2.4). This enzyme converts acetate and ATP to acetyl-CoA, AMP, and pyrophosphate. Consistent with this mechanism, the expression of *ACS1* was increased during growth on YPA as compared to YPG. In addition, we observed significantly higher expression of isocitrate lyase (*ICL1*, 154-fold) and malate synthase (*MLS1*, 24-fold). Both these enzymes are involved in the glyoxylate cycle and are predicted to reside in the cytosol. While acetate can cross the plasma membrane, many organisms also utilize transporters for acetate uptake [68]. The expression of two

predicted permeases (protein ID: 11570 and 10804) was increased almost 150 and 145-fold, respectively, during growth on YPA versus YPG. We also observed increased expression of acetyl-CoA hydrolase (*ACH1*). Increased expression of this enzyme is consistent with a detoxification mechanism.

We also observed increased expression of the genes involved in glycolysis, the pentose phosphate pathway, the TCA cycle, and the electron transport chain during growth on YPA versus YPG (Fig. 2.5). Fructose-1,6-biphosphatase (*FBP1*) and phosphoenolpyruvate carboxylase (*PCK1*) were both upregulated during growth on growth, suggesting increased flux through gluconeogenesis. In addition, we observed reduced expression of the genes involved in lipid production. Fatty acid biosynthesis genes including acetyl-CoA carboxylase (*ACC1*), and fatty-acyl-CoA synthase (*FAS1* and *FAS2*), were downregulated. Interestingly, the expression of ATP-citrate lyase (*ACL1*) also decreased. Likely, reduced expression of *ACL1* reflects the high concentration of cytoplasmic acetyl-CoA arising from acetate as the main carbon source. However, pathways involved in carbon storage and triacylglycerol (TAG) synthesis, including the glucan branching enzyme (*GLG3*), glycogen synthase (*GSY1*), myo-inositol-1-phosphate synthase (*INO1*), diacylglycerol *O*-acyltransferase (*DGA1*), acylglycerone-phosphate reductase (*AYR1*), glycerol 3-phosphate dehydrogenase (*GPD1*, *GPD2* and *GUT1*) were upregulated. Taken together, these results indicate that the genes involved in acetate utilization, including the glyoxylate cycle, are induced by acetate. Interestingly, we also found that the genes involved in central metabolism, namely glycolysis, were also upregulated.

During growth on YPA, we did not observe many significant changes in the concentration of intracellular metabolites as compared to growth on YPG. We mostly observed increased concentrations for several amino acids: alanine, tyramine, and threonine (Fig. 2.3). Only the concentration of lysine decreased. In agreement with the increased expression of carbon storage

pathways, we observed an increase in the concentration of glycerol, mannitol, and myo-inositol, and a decrease in the concentration of glycerol-3-phosphate. Lastly, the concentrations of intracellular long-chain fatty acids decreased, in alignment with the reduced expression of fatty acid biosynthesis.

Growth on oil

We lastly explored fatty-acid metabolism in *R. toruloides* by comparing gene expression during growth on YP medium containing soybean oil (YPS) versus YPG medium. Overall, 1296 genes were significantly upregulated, and 1392 genes were significantly downregulated during growth on YPS versus YPG.

Lipases catalyze the hydrolysis of fats and oils. *R. toruloides* has 10 predicated lipases. During growth on YPS, expression of two secreted lipases (*sLIP1* and *sLIP2*) and one intracellular lipase (*TGL1*), predicted to reside in the peroxisome, increased. Interestingly, the expression of the secreted sLIP3 lipase decreased, which hydrolyses triacylglycerol into free fatty acids and glycerol [69].

Not surprisingly, we observed increased expression of the genes involved in fatty-acid catabolism and decreased expression of the genes involved in fatty-acid synthesis, when comparing growth on YPS versus YPG. Fatty-acid ligases (FAA) are involved in the activation of imported long-chain fatty acids (C12–C18) by forming the associated acyl-CoAs [70,71]. Expression of *FAA1-5* was increased during growth on YPS as compared to YPG. *FAA2* is involved in the activation of medium-chain fatty acids directed towards peroxisomal β-oxidation, whereas *FAA3* is more active toward fatty acids with chain lengths longer than 18 [70]. We also observed increased expression of the *FAT1* fatty-acid transporter.

Fatty-acid catabolism is a multistep process and requires four different types of enzymes [72]. In the peroxisome, the first reaction is catalyzed by acyl-coenzyme A oxidase (ACOX) and acyl-CoA dehydrogenase (ACAD), which converts acyl-CoA to enoyl-CoA. *R. toruloides* has three ACOX's (*ACOX1-3*) and six different ACAD's (*ACAD1-6*) [41]. Multiple ACOX's isozymes in a single organism often have different specificities, including ones active against short-chain fatty acids, long-chain fatty acids, or both [73,74]. Expression of *ACOX1* and *ACOX3* was increased during growth on YPS. Among the six ACAD's, expression of *ACAD3* and *ACAD4* was significantly increased (>15 fold). The expression of the other three *ACAD*'s, *ACAD2*, *ACAD5* and *ACAD6*, were more moderately increased (~9-fold). The second and third steps are catalyzed by enoyl-CoA hydratase (ECH) and 3-hydroxyacyl-CoA dehydrogenase (HADH), which converts trans-enoyl-CoA to β-keto-acyl-CoA. The expression of both ECHs (*ECH1* and *ECH2*) and *HADH1* were increased. The expression of *FOX2*, the peroxisomal multifunctional enzyme, which catalyzes both steps 2 and 3 [75], was also increased. In the fourth step, 3-ketoacyl-CoA thiolase (POT1) catalyzes the thiolytic cleavage of the 3-ketoacyl-CoA intermediate, yielding one acetyl-CoA and an acyl-CoA molecule shortened by two carbon atoms [76]. We observed an almost 10-fold increase in the expression of *POT1*. We also observed increased expression of the genes involved in the glyoxylate cycle: malate synthase (*MLS1*), isocitrate dehydrogenase (*IDP1*) and isocitrate lyase (*ICL1*).

The peroxisome membrane is impermeable to the acyl-CoA's generated during the oxidation of long-chain fatty acids. The carnitine shuttle uses carnitine to channel shorter acyl groups from coenzyme A to form acetyl carnitine, which can be shuttled across membranes of the peroxisomes and mitochondria [77–79]. Carnitine acetyltransferase (CAT2), localized in the peroxisome, transfers an acyl group from acyl-CoA to carnitine. We observed increased expression of *CAT2* during growth on YPS versus YPG. Acyl-carnitine is then transported into

mitochondria by acylcarnitine translocase (CRC) [80]. We did not observe significant upregulation of *CRC*. The reverse transfer of the acyl group from acyl-carnitine to coenzyme A is catalyzed by mitochondrial *CAT2*, whose expression was also increased. This process regenerates acyl-CoA and carnitine inside the mitochondria.

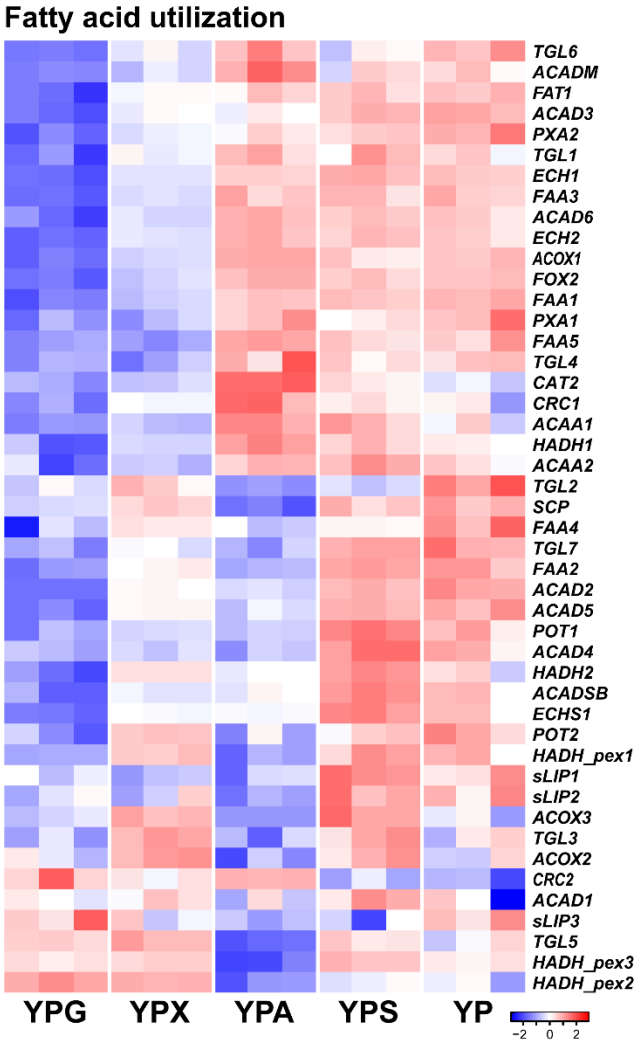

Figure 2.6. Transcriptional changes in pathways associated with fatty acid utilization in *R. toruloides* IFO0880 grown on glucose (YPG), xylose (YPX), acetate (YPA), soybean oil (YPS) and yeast peptone (YP). RNAseq data was collected in triplicate for each condition. Color key represents the z-score for each gene (normalized for all growth conditions).

Short-chain acyl-CoA that are carried through the carnitine shuttle, or activated in the cytosol by FAA2, enter the mitochondria for further oxidation. The mitochondrial oxidation also comprises of four enzymatic steps, similar to peroxisomal β-oxidation [77–79]. *R. toruloides* has mammalian orthologs of mitochondrial short chain and branched chain acyl-CoA dehydrogenases (ACADSB) and medium chain acyl-CoA dehydrogenases (ACADM). The first reaction is catalyzed by ACADSB and ACADM, which converts long and short chain acyl-CoA's to trans-enoyl-CoA. We observed a significant (50-fold) increase in the expression of *ACADSB. ACADM* expression was also increased, albeit more moderately (2-fold). We also observed increased expression of enoyl-CoA hydratase (*ECHS1*) and 3-hydroxyacyl-CoA dehydrogenase (*HADH1* and *HADH2*), which convert trans-enoyl-CoA to β-keto-acyl-CoA. Finally, the expression of acetyl-CoA acetyltransferase (*ACAA1* and *ACAA2*), which catalyze the conversion of β-keto-acyl-CoA into one acetyl-CoA and an acyl-CoA molecule shortened by two carbon atoms, was also increased. Overall, these changes are consistent with the utilization of oil.

Finally, we observed increased expression of many genes involved in the TCA and glyoxylate cycle. Increased expression of the glycolytic enzymes, phosphofructokinase (*PFK1/2*), pyruvate carboxylase (*PYC1*), phosphoenolpyruvate carboxykinase (*PCK1*), and hexokinase (*HXK*), was also observed. However, expression of pyruvate kinase (*PYK1*), pyruvate dehydrogenase (*PDB1*) and glyceraldehyde-3-phosphate dehydrogenase (*GADPH*) was decreased. Since the respiratory chain and TCA cycle are metabolically connected with mitochondrial β-oxidation, these changes in expression are likely due to increased β-oxidation.

When comparing intracellular metabolites during growth on YPS versus YPG, we observed increased concentrations of mannitol and xylitol (Fig. 2.3). The origin for these differences is not immediately evident. Interestingly, we observed a reduction in the concentration of many fatty-acid derived molecules, which likely reflects the fact that the cells are degrading fatty acids

rather than making them. In addition, both the concentrations of ornithine and ornithine-1,5-lactam decreased. Likely, the cells are using fatty acids for energy and the amino acids in YPS medium for biomass, unlike the case during growth on YPG.

2.4 Discussion

R. toruloides can grow on multiple carbon sources, including glucose, xylose, acetate, galactose, arabinose, mannose, fructose, sucrose, and oil. However, little is known about the governing metabolic pathways and their regulation. To better understand these pathways, we measured gene expression and intracellular metabolites during growth of *R. toruloides* IFO0880 on glucose, xylose, acetate, and oil. These substrates were chosen because they are being explored as potential substrates for conversion by *R. toruloides* into more valuable fuels and chemicals. All the samples were collected at exponential phase of YPG, YPX, YPA, and YPS medium in a baffled shake flask. The sampling point for the YP control was late exponential. *R. toruloides* can grow well on YP medium without additional carbon source. We selected the 24 h sampling time point for the YP control to ensure a sufficiently high OD and also used a similar time as the glucose, xylose, and acetate experiments. The sample point for oil was chosen earlier because it already reached an OD of 10 after 16 h of growth. Significant differences were observed in gene expression and metabolite concentrations during growth on the different substrates (Fig. 2.2 and 2.3). To better understand these differences, we mapped them onto the central carbon metabolic pathways of *R. toruloides* (Fig. 2.4). This analysis required us to identify and annotate many of the genes governing these pathways.

Most of the changes in gene expression and metabolite concentrations were consistent with the metabolism of these substrates. During growth on glucose, expression of the genes involved in lower glycolysis and the TCA cycle was increased. Expression of *GDH1* was also increased by more than 100- fold. GHD1 catalyzes the reversible conversion of ammonia and α-

ketoglutarate to form glutamate using NADPH as the cofactor. Likely, the large increase in *GDH1* expression reflects the need of *R. toruloides* to scavenge nitrogen from amino acids in the growth medium. With respect to intracellular metabolites, the concentrations of glycolytic metabolites, galactose, and ornithine were higher during growth on YPG versus YP. Glucose can be converted to galactose, which is likely used to synthesize polysaccharides, glycolipids, and glycoproteins. In addition, the concentration of ornithine, which is involved in the urea cycle, increased. This likely reflects increased utilization of amino acids for production of energy during growth on glucose.

Like many yeasts, *R. toruloides* employs the xylose reductase (XR)/xylitol dehydrogenase (XDH) pathway for xylose utilization (Fig. 2.4). The first enzyme, xylose reductase, converts xylose to xylitol using NADPH as the cofactor. The second enzyme, xylitol dehydrogenase, converts xylitol to xylulose using NAD as the cofactor. In most yeasts, xylulokinase (XKS) converts xylulose to xylulose-5-phosphate, which then enters the non-oxidative branch of the pentose phosphate pathway. Growth on xylose principally increases the expression of genes involved in the non-canonical xylose utilization pathway in *R. toruloides* and has only a minor effect on other genes involved in central metabolism, though it does affect the expression of many other genes. Growth on xylose resulted in increased expression of the genes involved in xylose metabolism, with the notable exception of xylulokinase. This suggest that there may be xylose-specific transcriptional regulators. However, the identity of these regulators and associate processes are unknown.

One interesting finding concerns xylose metabolism. Previously, *R. toruloides* was found to produce relatively large amounts of arabitol during growth on xylose [64]. This behavior is puzzling because it does not have any obvious benefit to the cell. Our analysis suggests that this behavior is due to poor expression of xylulokinase, resulting in a bypass through arabitol

dehydrogenase. Even stranger is that xylose reduces xylulokinase expression. One possibility is that this is a strain specific mutation and unique to *R. toruloides* IFO0880. However, many yeasts produce arabitol during growth on xylose, suggesting that it is an adaptive strategy of unknown benefit [81,82]. Regardless, these results provide a potential strategy for improving the growth on *R. toruloides* on xylose, which natively is very poor, by constitutively expressing xylulokinase.

In the case of acetate, expression of the genes involved in TCA cycle, gluconeogenesis, and pentose phosphate pathway were increased. Acetate enters the central metabolism through the action of acetyl-CoA synthetase (Fig. 2.4). This enzyme converts acetate and ATP to acetyl-CoA, AMP, and pyrophosphate. Increased expression of two permeases (protein ID: 11570 and 10804) suggest that they may facilitate acetate uptake. However, more detailed studies are required before proving this claim. Collectively, these results are consistent with acetate inducing the expression of acetate utilization genes. Acetyl-CoA hydrolase converts acetyl-CoA into acetate. However, increased expression of this enzyme would result in a futile cycle that consumes ATP by hydrolyzing the acetyl-CoA formed by acetyl-CoA synthetase. Researchers have previously shown that this enzyme in many fungi is a CoA-transferase involved in the detoxification of intracellular acetate by transforming acetate and succinyl-CoA into acetyl-CoA and succinate [83,84].

Lastly, during growth on soybean oil, we observed increased expression of many secreted lipases and putative fatty-acid transporters. Fatty acid transporter (FAT1) has been reported to play a pivotal role in the import of long-chain fatty acids in *Saccharomyces cerevisiae* and *Yarrowia lipolytica* [85,86]. In another study, it was reported that FAT1 is another member of the yeast long-chain acyl-CoA synthetase family [87,88]. In vitro analysis has shown that FAT1 has CoA ligase activity towards very long-chain fatty acids (C22-C26) [89]. These results suggest that,

in addition to transporter activity, FAT1 can also activate long-chain fatty acids (LCFAs) by forming the associated acyl-CoA's.

Following the import aided by FAA's and FAT1, fatty acids are degraded by β-oxidation. We observed increased expression of the genes involved in β-oxidation in both the mitochondria and peroxisomes. In *R. toruloides*, both mitochondrial and peroxisomal fatty acid β-oxidation pathways have been reported [13,41]. Long-chain fatty acids are shortened in the peroxisome and then transferred via the carnitine shuttle to the mitochondria, where short chain fatty acids are oxidized to acetyl-CoA. Both peroxisomal and mitochondrial beta-oxidation are necessary for robust growth on fatty acids [90,91].

In conclusion, these results further our understanding of central carbon metabolism in *R. toruloides.* These results and the associated data sets may help metabolic engineers further exploit the ability of *R. toruloides* to make diverse fuels and chemicals. In addition, they may also aid in the development of genome-scale models [26,43,92,93]. Future work is focused on understanding the mechanisms governing lipid production during nutrient starvation and identifying the regulatory mechanisms governing metabolism.

Chapter 3: Gene Regulation of Substrate Utilization in *Lipomyces Starkeyi*

3.1 Introduction

Oleaginous yeasts have been studied extensively due to their substantial lipid storage capability [94,95]. The accumulated lipids provide alternatives to plant oils for biodiesel production [96]. *Lipomyces starkeyi* is a promising oleaginous yeast, which was isolated from soil by Starkey [97] and later described by Lodder and Kreger-Van Rij [98]. It can grow on various plant-based sugars and is an excellent lipid producer [99]. Because of its ability to grow on a wide variety of sugars, many enzymes like dextranases, amylase, levoglucosan kinase, and other hydrolases have been studied and characterized from *L. starkeyi* [100–105].

Nutrient limitation in *L. starkeyi* has been reported as a mechanism to induce lipid production [106,107]. Recent work has also explored the pathways involved in lipid metabolism [108,109]. In addition to lipids, *L. starkeyi* has been engineered for fatty alcohol production [110,111]. Lipid production in *L starkeyi* has been optimized using two-stage fermentation for industrial use [112–114]. In addition to accumulating large amounts of lipids, *L. starkeyi* can naturally utilize the sugars present in lignocellulosic hydrolysates and is tolerant to the inhibitors present in these hydrolysates [115,116]. However, there is a considerable lack in our understanding of *L. strkeyi*, primarily due to insufficient knowledge of its physiology and the lack of efficient genetic tools.

The genome sequence of *L. starkeyi* NRRL Y-11557 was reported in 2016, along with 15 other biotechnologically relevant yeasts [117]. Since then, a few studies have used the gene models of *L. starkeyi* NRRL Y-11557 in different bioinformatics analyses. Theoretical lipid yields on various sugars were reported using a small-scale metabolic model of *L. starkeyi* NRRL Y-11557 [118]. A recent study reported the transcriptional changes in *L. starkeyi* NRRL Y-11558 resulting from growth on hydrolysate [116]. In addition, two studies have used bioinformatics analysis to

identify and characterize the sugar transporters present in *L. starkeyi*: LST1_120451 as a cellobiose transporter [119] and LST1_205437 as a co-transporter for glucose and xylose [120]. Another study constructed the regulatory network of TAG biosynthesis in *L. starkeyi* using metabolic profiling under nitrogen limiting conditions [121]. Multiple transformation protocols have been established for the genetic engineering of *L. starkeyi* [122–125], and heterologous gene expression has been realized [111,126], enabling *L. starkeyi* to serve as a promising oleaginous yeast for chemical and fuel production.

In this work, we have systematically evaluated the growth of *L. starkeyi* NRRL Y-11557 on the plant-based sugars: glucose, xylose, and cellobiose. We performed transcriptomic and metabolomic analyses to better understand the underlying mechanisms of sugar uptake by this yeast. This work reports the different pathways activated during growth on these sugars based on the genome-wide gene expression differences.

3.2 Materials and methods

Strains, media, and culture conditions

Lipomyces starkeyi NRRL Y-11557 (DSM 70295; sourced from DSMZ – German Collection of Microorganisms and Cell Cultures) was grown on YPD medium (10 g/L yeast extract, 20 g/L peptone, and 20 g/L glucose) at 30 °C for routine culture. YPX medium (10 g/L yeast extract, 20 g/L peptone, and 20 g/L xylose), YPC medium (10 g/L yeast extract, 20 g/L peptone, and 20 g/L cellobiose), and YPD medium were used for transcriptomic and metabolomic analysis.

HPLC analysis of extracellular metabolites

Yeast growth (OD600) was measured using a spectrophotometer (Biomate 5, Thermo, NY). Extracellular metabolites such as glucose, xylose, cellobiose, glycerol, acetate, and ethanol were measured by HPLC (Agilent Technologies 1200 Series, Santa Clara, CA) with a RezexTMROA-

Organic Acid H + (8%) column (Phenomenex Inc., Torrance, CA) and a refractive index detector (RID). The column was eluted with 0.005 N H_2SO_4 at a 0.6 mL/min flow rate at 50 °C.

RNA sequencing analysis

A single colony from a YPD agar plate was inoculated into YPD liquid medium to obtain *L. starkeyi* seed cultures. Seed cultures were then used to inoculate 50 mL YPD, YPX, and YPC medium in 250 mL baffled shake flasks, with a starting OD600 of 1. The cells were then grown at 250 rpm, 30 °C. The growth was monitored by measuring OD600 and sugar concentration. The samples for transcriptomics analysis were taken during mid-exponential growth, as shown in Fig. 3.1 (30 h for YPD, YPX, and 52 h for YPC). The samples were collected and washed using cold ddH_2O. Total RNA was extracted using RNeasy Mini Kit (Qiagen, Hilden, Germany) and then treated with DNA-free DNase using the TURBO DNA-free kit (Ambion, Austin, TX) to remove genomic DNA. The quality of RNA was confirmed by agrase gel electrophoresis and bioanalyser. DNA gel and Nanodrop. The stranded RNAseq libraries were prepared with Illumina's TruSeq Stranded RNA Sample Prep kit. The libraries were pooled in equimolar concentration and sequenced for 101 cycles from each single end of the fragments on a HiSeq2500 (Illumina). Fastq files were generated and demultiplexed with the bcl2fastq v1.8.4 Conversion Software (Illumina, San Diego, CA, USA).Raw sequencing reads are available at NCBI (BioProject ID: PRJNA808049).

RNA sequencing resulted in 9 samples with 13-19 million reads. The sequencing analysis was conducted using an in-house pipeline, as previously described [127,128]. Adaptor sequences and low-quality reads were trimmed and analyzed using Trimmomatic [45] and FastQC [129]. Reads were mapped to the *L. starkeyi* NRRL Y-11557 reference genome (NCBI GenBank assembly accession: GCA_001661325.1) [117] with STAR version 2.5.3a [47]. Read counts were calculated

using featureCounts from the Subread package, version 1.5.2 [48]. Differential expression analysis was performed on the reads counts in R v4.0.5 using edgeR v3.32.1 and limma v3.46.0 [49,50]. Graphical representation of expression data was constructed using R packages: PCAtools v2.2.0 [51], gplots v3.1.1 [53], and Glimma v2.0.0 [52]. Functional annotation of *L. starkeyi* was obtained from DOE Joint Genome Institute website [54]. Data analysis scripts, along with the results, can be downloaded from the following link: https://github.com/raogroupuiuc/lipo11557_growth.

Pairwise enrichment analysis was performed using GSEA [130]. Briefly, the KEGG and GO annotation files downloaded from JGI Mycocosm were reformatted to the GSEA guidelines using Python. All parameters were set to default in GSEA, except 'Permutation Type' was set to 'gene_set'. Enrichment analysis was performed separately for different annotation classes (GO: biological process, molecular function, and cellular component and KEGG: pathways, and pathway class) and visualized using Cytoscape v3.8.2 [131]. DeepLoc-1.0 was used to predict the subcellular localization of proteins [132].

Metabolomics analysis

Yeast cells grown in YPD, YPX, and YPC were collected during exponential phase and vacuum filtered using Vac-Man Laboratory Vacuum Manifold (Promega, Madison, WI) assembled with a nylon membrane filter (pore size, 0.45 μm; diameter, 13 mm; Whatman, Piscataway, NJ) and a filter holder (Millipore, Billerica, MA). The filtered cell culture was washed with 2.5 mL of prechilled distilled water. The entire process of fast filtration was completed within 1 min as previously described [127]. The filter membrane containing the washed cells was quickly mixed with 1 mL prechilled acetonitrile-water mixture (1:1, v/v) and 100 μL of glass beads. The mixture was vortexed for 3 min to disrupt cell membranes, allowing extraction of intracellular metabolites. The extraction mixture was then centrifuged at 16,100 × *g* for 3 min

at 4 °C, and 0.8 mL of the supernatant containing the intracellular metabolites was dried in a speed vacuum concentrator for 6 h.

Before GC/MS analysis, the samples were derivatized by methoxyamination and trimethylsilylation, as previously described [127]. For GC/MS, the derivatized metabolite samples were applied to an Agilent 7890A GC/5975C MSD system (Agilent Technologies) equipped with an RTX-5Sil MS capillary column (30 m × 0.25 mm, 0.25 μm film thickness: Restek, Bellefonte, PA) and an additional 10-m-long integrated guard column. One microliter of the derivatized sample was injected into the GC inlet in splitless mode. The oven temperature was initially set to 150 °C for 1 min, after which the temperature was increased to 330 °C at 20 °C/min, where it was held for 5 min. The mass spectra were recorded in a scan range 85–500 m/z at an electron impact of 70 eV, and the temperatures of the ion source and transfer line were 230 and 280 °C, respectively.

The raw data obtained from the GC-MS analysis were processed using an automated mass spectral deconvolution and identification system (AMDIS) software for peak detection and deconvolution of mass spectra. The processed data were uploaded to SpectConnect (http://spectconnect.mit.edu) for peak alignment and generation of the data matrix with the Golm Metabolome Database mass spectral reference library. The normalized abundance values for each metabolite were obtained by dividing peak intensity with dry cell weight. For statistical analysis, such as principal component analysis (PCA) and clustering analysis (represented as a heatmap), Statistica (version 7.1; StatSoft, Tulsa, OK, USA), MetaboAnalyst, and MultiExperiment Viewer software were used [60,61]. In addition, GraphPad Prism 6 (GraphPad, San Diego, CA, USA) was used for plotting bar graphs.

3.3 Results

Growth and utilization of different substrates

The hydrolysis of lignocellulosic biomass releases monosaccharide and disaccharide sugars, principally in the form of glucose, xylose, and cellobiose. *L. starkeyi* is an oleaginous yeast capable of utilizing these lignocellulosic sugars. In this work, we test the growth of *L. starkeyi* NRRL Y-11557 on glucose, xylose, and cellobiose. *L. starkeyi* was grown on YP medium containing one of the following sugars: glucose (YPD), xylose (YPX), and cellobiose (YPC). Fig. 3.1 shows the growth and substrate utilization of *L. starkeyi* on the different sugars. Growth on the three sugars led to similar final cell densities. However, we observed differences in sugar utilization. The cells ultimately utilized 20 g/L glucose and xylose in 48 h, whereas consumption of 20 g/L cellobiose was slower and took 84 h of growth. *L. starkeyi* consumed glucose marginally faster than xylose, which is evident from the timepoint for 50% sugar utilization. The cells consumed 10 g/L of glucose, xylose, and cellobiose in 28 h, 32 h, and 56 h respectively.

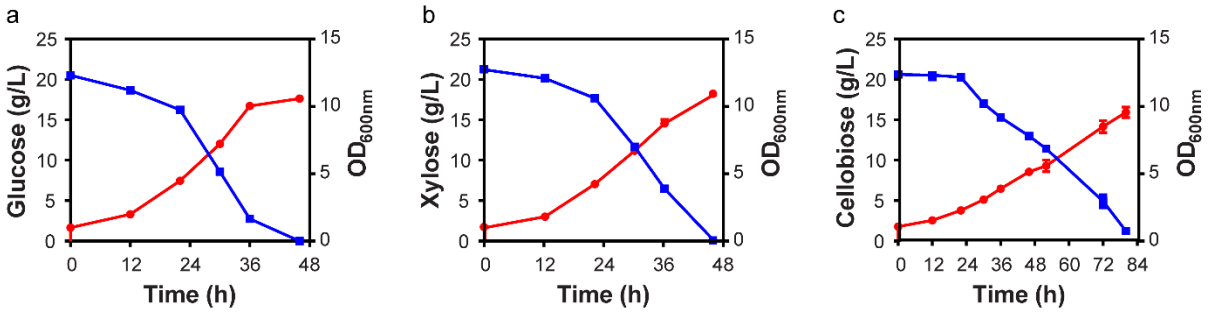

Figure 3.1. Growth profiles of *L. starkeyi* NRRL Y-11557. Growth on 20 g/L of glucose (a), xylose (b), and cellobiose (c) in rich medium. Sugar concentration is denoted using the blue squares and plotted on the left y-axis. Optical density (OD_{600}) is denoted using the red circles and plotted on the right y-axis.

Identification of genes associated with substrate utilization

We used whole-genome RNA sequencing to analyze the growth of *L. starkeyi* on the three sugars. RNA was extracted approximately at the time point for 50% sugar utilization, with 3

biological replicates for each condition (30 h for glucose and xylose, and 52 h for cellobiose). We chose the mid-point of sugar utilization in each case to capture the gene expression profiles for sugar uptake and metabolism. The sequencing resulted in an average of 16.3 million reads across 9 samples. More than 85% of reads were mapped to a unique location in the reference genome (NCBI GenBank assembly accession: GCA_001661325.1) [117]. The different clusters on the principal component analysis (PCA) plot show the distinct gene expression profiles of the three sugars (Fig. 3.2a). PC1 axis, representing 70% variation, highlights the difference between growth on glucose and cellobiose, whereas PC2 axis, which represents 22% variation, distinguished the growth on xylose from the other two sugars.

We chose glucose as a control to compare the differences in gene expression profiles of different sugars for further analysis. During the growth on cellobiose compared to glucose, expression of 851 genes was significantly increased, and expression of another 374 genes was significantly decreased. During growth on xylose compared to glucose, expression of 392 genes was significantly increased and 120 genes was significantly decreased. Fold change > 2 and adjusted *p*-value < 0.05 were considered significant for comparative gene expression analysis. These numbers are in-line with the clustering trend on the PCA plot.

To further investigate the differences in global gene expression profiles of these sugars, we performed gene set enrichment analysis (GSEA) using KEGG annotations. Differential gene expression during growth on cellobiose is limited to sugar uptake and metabolism, like starch metabolism, pentose and glucuronate interconversion pathway, and galactose metabolism (Fig. 3.2b). On the other hand, growth on xylose resulted in broader changes in the central carbon metabolism, as noted in Fig. 3.2c. We note that while cellobiose has a more extensive set of differentially expressed genes than xylose, the list of enriched pathways is smaller than xylose.

(Fig. 3.2b, c). This is potentially because we have limited information about the gene annotation of those gene sets.

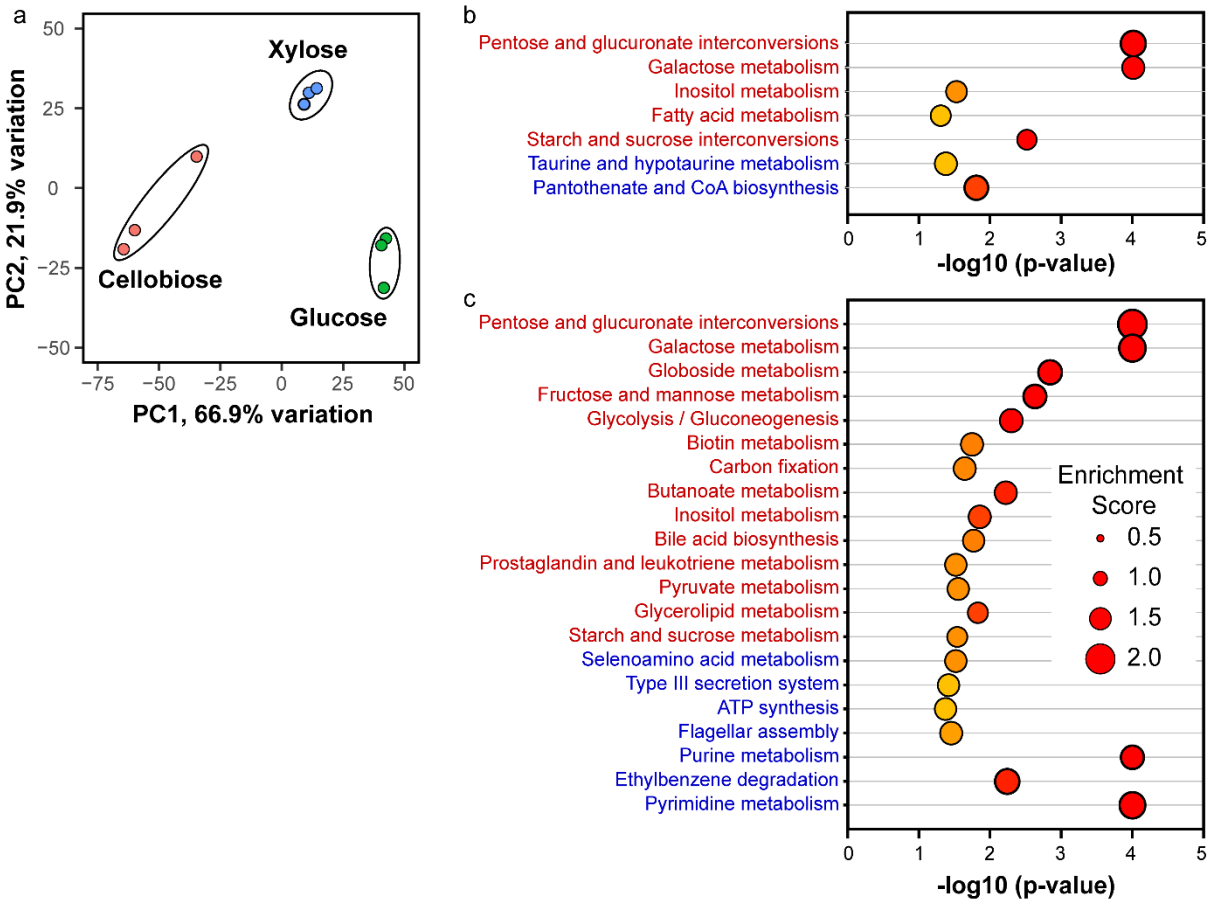

Figure 3.2. Overview of the transcriptomics analysis. (a) Principal component analysis plot generated from gene expression profiles of *L. starkeyi* grown on glucose, xylose, and cellobiose. RNAseq data were collected in triplicates for each condition. (b and c) KEGG pathway enrichment analysis for growth on cellobiose vs. glucose (b) and xylose vs. glucose (c). The vertical axis represents the pathway category, and the horizontal axis represents the pathway's enrichment score [−log(p-value)]. Significantly enriched KEGG pathways ($p < 0.05$ and FDR < 0.25) are plotted. The data were analyzed using the GSEA tool and plotted in Origin.

Comparative analysis of growth on cellobiose

We first explored cellobiose metabolism in *L. starkeyi* by comparing the gene expression on YP medium containing cellobiose (YPC) versus YP medium containing glucose (YPD). RNA was extracted at the midpoint of sugar consumption for both conditions to capture the gene expression differences resulting from the two sugars. Overall, 851 genes were significantly

upregulated, and another 374 genes were downregulated during growth on cellobiose versus glucose.

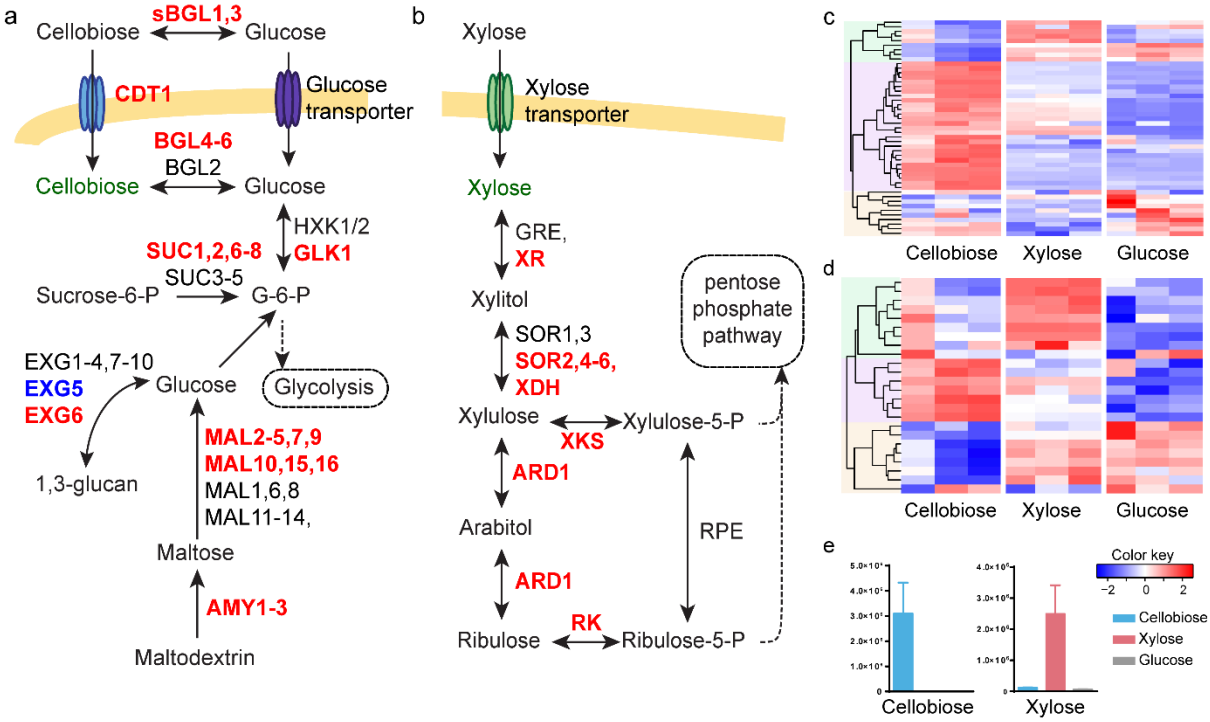

Figure 3.3. Substrate utilization in *L. starkeyi* (a and b). Substrate uptake and metabolism for cellobiose (a) and xylose (b). Differential gene expression is reported in comparison to glucose. Upregulated genes (fold change > 2 and p-value < 0.05) are highlighted in red and downregulated genes in blue. (c and d) Heatmap of gene expression data in KEGG pathways: starch and sucrose metabolism (c) and pentose and glucuronate metabolism (d). RNAseq data were collected in triplicate for each condition. Color key represents the z-score for each gene (normalized for all growth conditions). (e) Intensity levels of intracellular metabolites profiled using GC/MS during growth on different sugars. The data represents the average of six independent measurements, and bars indicate standard deviation.

We focused on the cellobiose uptake pathways in Fig. 3.4a. Cellobiose can be transported into and utilized by the cell in two ways. The first is the import of cellobiose with the aid of a cellobiose transporter and further hydrolysis of intracellular cellobiose into glucose by β-glucosidase. A recent study identified LST1_120451 as a cellobiose transporter (CDT1) [119].

Expresssion of the *CDT1* gene was increased 11-fold during growth on cellobiose versus glucose. *L. starkeyi* has multiple copies of the gene β-glucosidase (*BGL1-6*), with the increased expression of five during growth on cellobiose as compared to glucose. Expression of glucokinase (*GLK1*), the enzyme that facilities the phosphorylation of glucose, was increased 2.3-fold. We also used DeepLoc [132], a localization prediction tool for eukaryotic genes, to predict the localization of BGL1-6. We note that BGL1 and BLG3 are extracellular, BGL5 and

BGL6 are peroxisomal, whereas BGL2 and BGL4 are cytosolic. The increased expression of *BGL3* was the highest amongst the six *BGL*'s (1252-fold), which alludes to the second mechanism of cellobiose utilization. Cellobiose is hydrolyzed into glucose extracellularly, aided by secreted β-glucosidases, and the resultant glucose is transported into the cell using glucose transporters. Multiple membrane-localized genes, annotated as predicated transporters belonging to the major facilitator superfamily, exhibited increased expression on cellobiose. However, we lack experimental validation for these predicted transporters. We previously tested a few predicted transporters for glucose uptake and found that LST1_205437 facilitates both glucose and xylose uptake [120].

We also observed non-specific regulation in *L. starkeyi* during growth on cellobiose. Expression of hydrolase genes, involved in the breakdown of other oligosaccharides, was also increased. (Fig. 3.4a). Nine α-glucosidase genes (*MAL*), which are involved in the breakdown of maltose, and three α-amylase (*AMY*), which are responsible for the breakdown of maltodextrin, and five β-fructofuranosidases (*SUC*) exhibited increased expression. Significant increases in expression were observed for the genes belonging to the starch and sucrose metabolism pathway of the KEGG pathway database (Fig. 3.4c) [133].

We next focused on the genes involved in central carbon metabolism (Fig. 3.5). Cellobiose breaks down into glucose and enters glycolysis. Genes in glycolysis did not exhibit significant

differences in expression, except for enolase (*ENO1*), whose expression decreased 2.2-fold, and fructose bisphosphate (*FBP1*), whose expression increased 2.4-fold. We also noticed decreased expression of both the pyruvate decarboxylases (*PDC1:* 2.6-fold ,*PDC2:* 3.4-fold) and a 4.9-fold increase in expression of pyruvate carboxylase (*PYC2*).

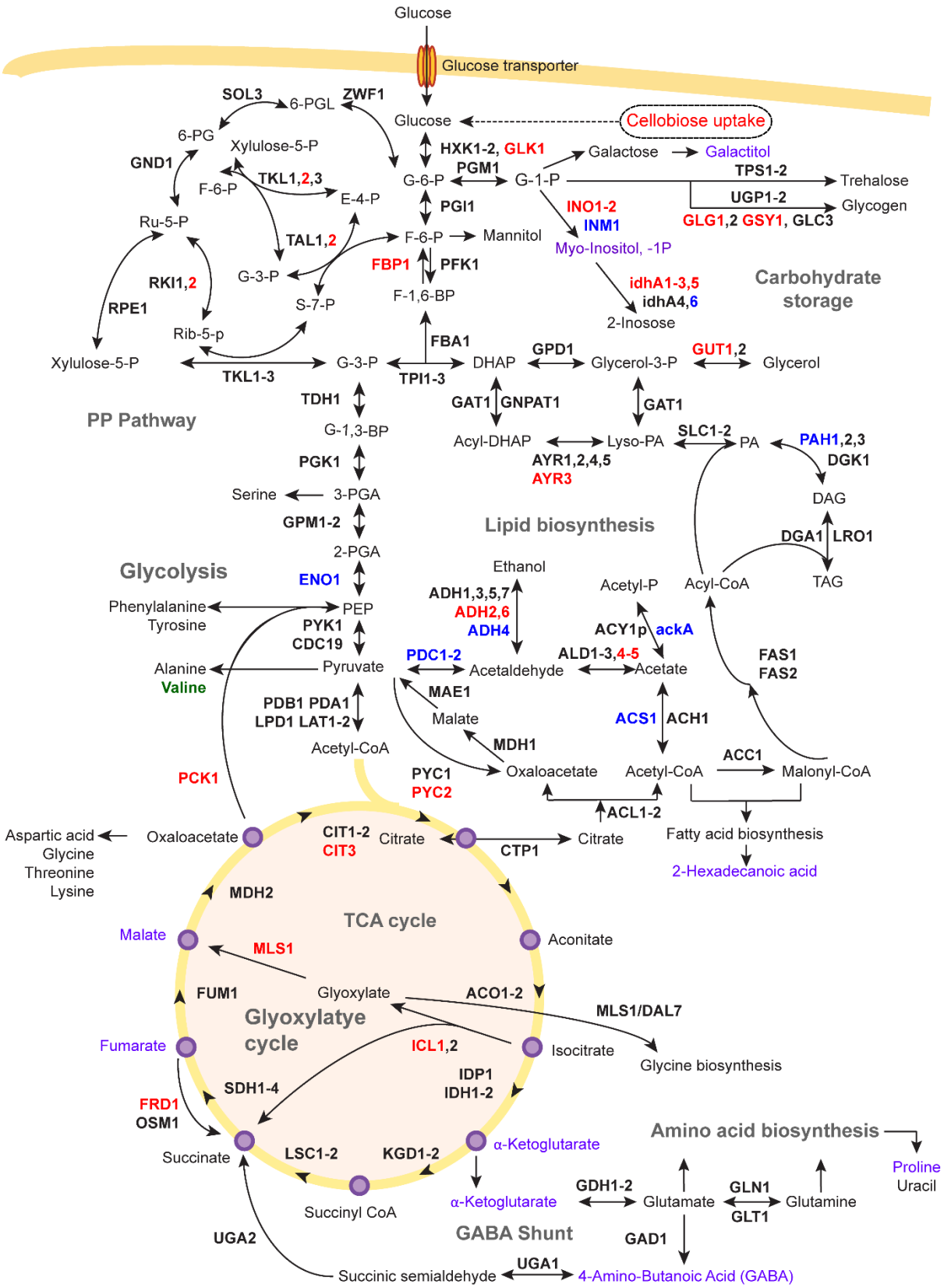

Figure 3.4. Differential gene expression of the central metabolic pathway during growth on cellobiose, reported in comparison to glucose.

Figure 3.4. (cont.) Upregulated genes are highlighted in red, and downregulated genes in blue. Metabolites with significantly higher abundance are highlighted in green, and metabolites in lower abundance in purple.

However, the small change in the expression of these genes indicates the absence of any global shifts in regulation induced from growth on cellobiose. One interesting result was the increased expression of multiple aldehyde dehydrogenases (*ALD1*: 20-fold, *ALD4*: 3-fold, and *ALD5*: 10-fold). We also observed decreased expression of acetate kinase (*ackA*) and acetyl-CoA synthetase (*ACS1*). It is plausible that *L. starkeyi* accumulates a smaller pool of acetyl-CoA during growth on cellobiose compared to glucose.

In the TCA cycle, we observed increased expression of one of the three citrate synthases (*CIT3*) by 7.9-fold and fumarate reductase (*FRD1*) by 14.3-fold. In addition, expression of both genes in the glyoxylate bypass, malate synthase (*MLS1*) and mitochondrial isocitrate lyase (*ICL1*), was increased 3.4-fold and 7-fold, respectively.

Differential analysis of growth on xylose

We next analyzed xylose metabolism in *L. starkeyi*. Gene expression during growth on YP medium containing xylose (YPX) is compared to YPD. Overall, 392 genes were significantly upregulated, and 120 genes were significantly downregulated during growth on xylose compared to glucose. The growth on xylose had a closer gene expression profile to glucose, with key differences in carbohydrate metabolism (Fig. 3.2c).

A prominent pathway for xylose utilization in eukaryotes is the oxidoreductive pathway, also known as the XR-XDH pathway (Fig. 3.4b) [134]. Xylose reductase (XR) reduces intracellular xylose to xylitol, which is further oxidized to xylulose by xylitol dehydrogenase (XDH) or sorbitol dehydrogenase (SOR1-6). The last step of the XR-XDH pathway is the phosphorylation of xylulose by xylulokinase (XKS), which enters the pentose phosphate pathway. In *L. starkeyi*,

we found that expression of all genes in the XR-XDH pathway was increased. For instance, we noted significant increases in expression for *XR* and *SOR2*, which increased 30 and 302-fold, respectively. In addition, expression of *XKS* was 27-fold increased as well. Increased expression was also observed in the genes belonging to the pentose and

glucuronate interconversion pathway of the KEGG pathway database (Fig. 3.4d). Some oleaginous yeasts have an arabitol bypass [64], where the carbon flux from xylose is directed through arabinitol dehydrogenase (ARD1) and ribulokinase (RK) instead of the one-step XKS. Expression of *ARD1* and *RK* was increased 2.4 and 5.4-fold, respectively. The extent of upregulation of these pathways in *L. starkeyi* indicates that the one-step XKS phosphorylation draws a higher carbon flux than the arabitol bypass.

Xylose uptake pathway converts xylose to xylulose-5P and ribulose-5P, which enter the pentose phosphate pathway (Fig. 3.6). Few genes in pentose phosphate pathways exhibited increased expression in *L. starkeyi*, which is consistent with the mechanism of xylose metabolism. Most notably, expression of transketolase (*TKL2*), transaldolase (*TAL1,2*), and ribokinase (*RBK1*) was increased. We also observed increased expression of a few genes in upper glycolysis, namely glucose-6-phosphate isomerase (*PGI1*) and fructose bisphosphate (*FBP1*). Genes involved in glyoxylate bypass also exhibited increased expression during growth on xylose, which was also the case with cellobiose growth. We observed a few more similarities in the gene expression on xylose and cellobiose. For instance, expression of both fumarate reductase and citrate synthase were increased.

Measurement of intracellular metabolites

To supplement the RNA sequencing data, we used gas chromatography-mass spectrometry (GCMS) and measured changes in the concentration of intracellular metabolites. Of the 55

metabolites measured, significant differences were observed in the concentrations of 37 metabolites during growth on glucose, xylose, and cellobiose (Fig. 3.3).

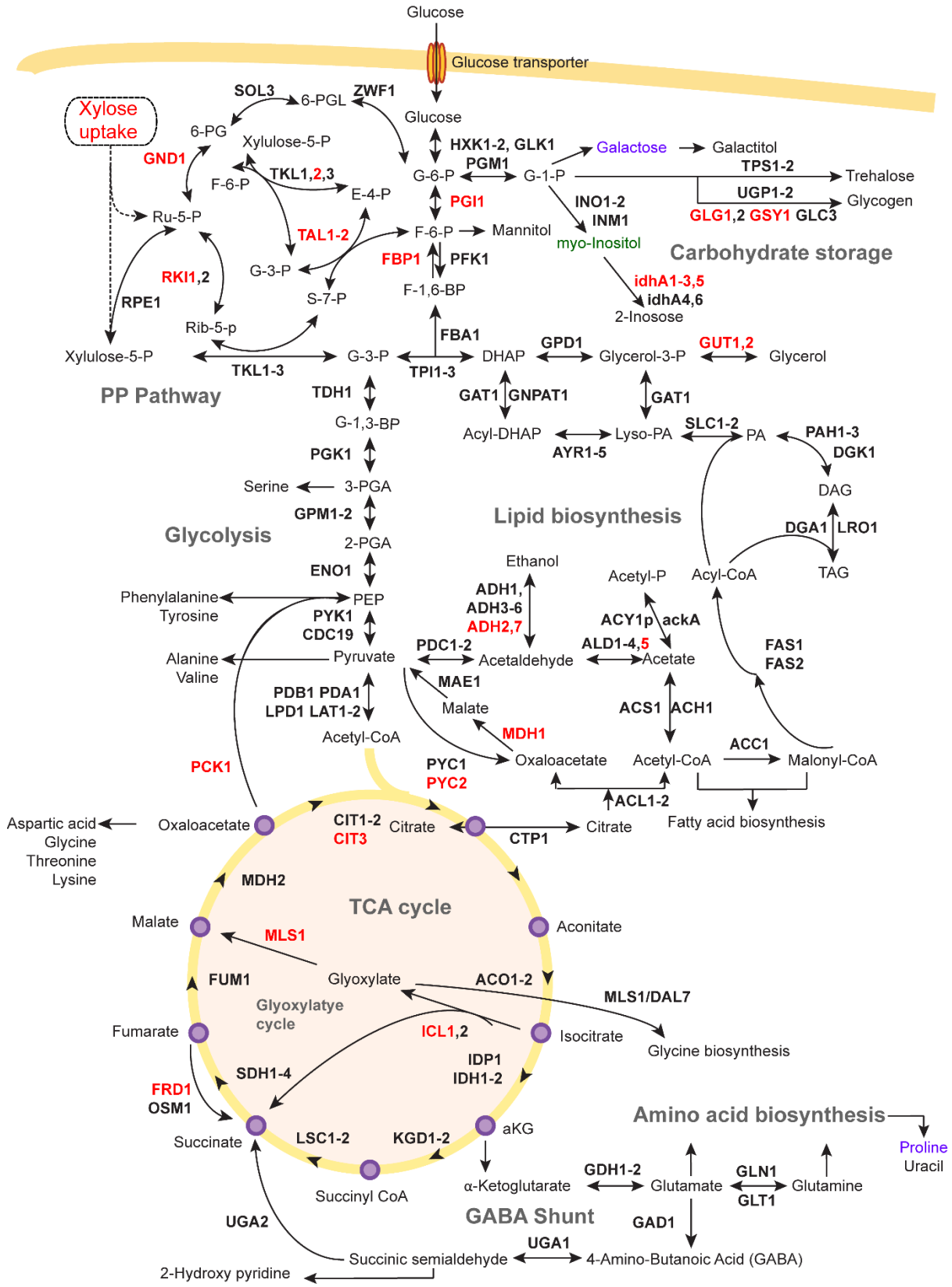

Figure 3.5. Differential gene expression of the central metabolic pathway during growth on xylose, reported in comparison to glucose.

Figure 3.5. (cont.) Upregulated genes are highlighted in red, and downregulated genes in blue. Metabolites with significantly higher abundance are highlighted in green, and metabolites with lower abundance in purple.

Significance was evaluated using two-sample t-tests, and the cut-off was relative metabolite concentration > 2 and adjusted *p*-value < 0.05. Compared to glucose, the concentration of 9 and 3 metabolites was higher on cellobiose and xylose, respectively. Whereas another 16 and 4 metabolites had a lower concentration in cellobiose and xylose, respectively, the changes in the metabolite concentrations align with those of differential gene expression.

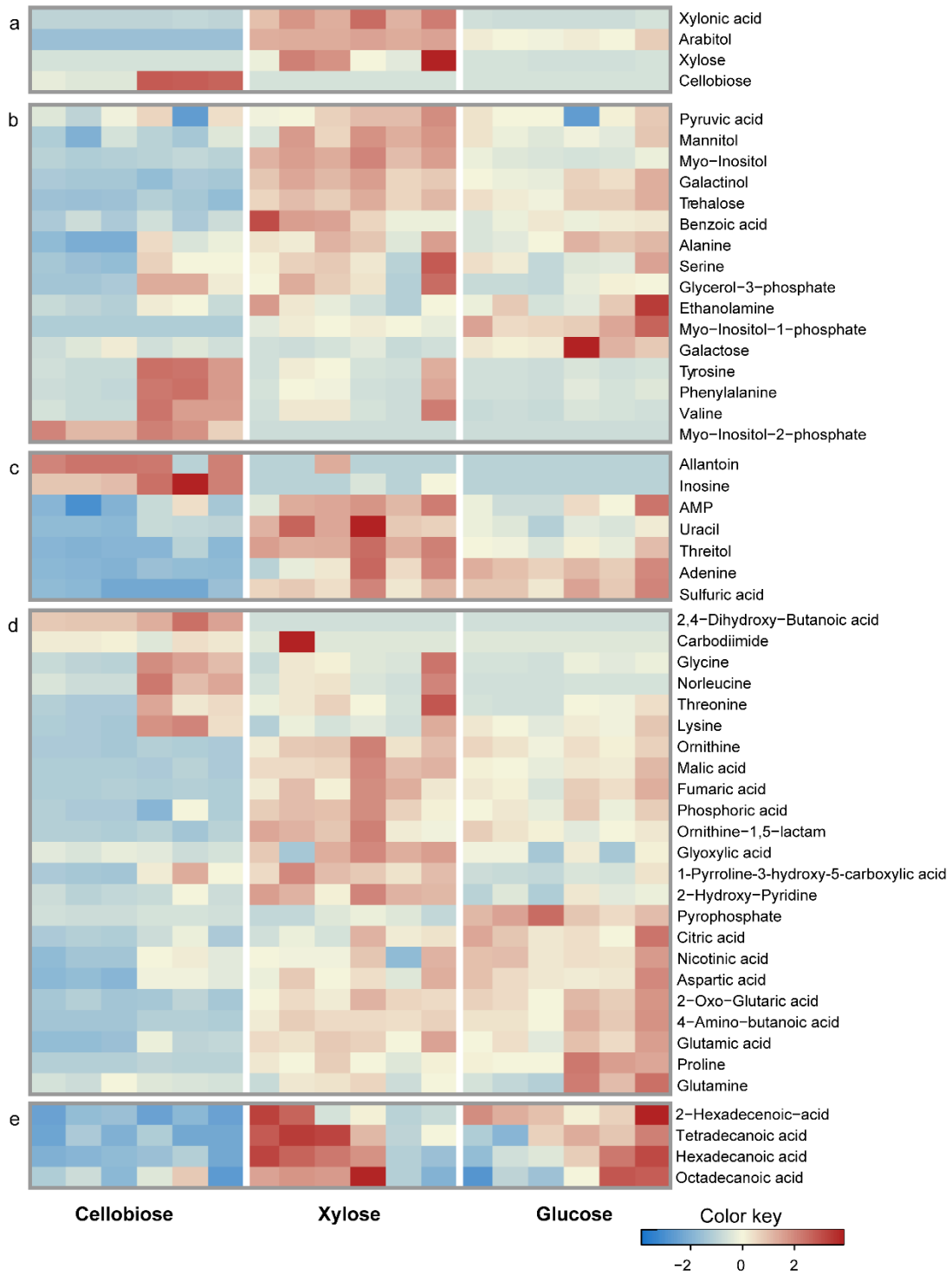

Figure 3.6. Overview of the metabolomics analysis. Heatmap of intracellular metabolites from xylose and cellobiose uptake pathways (a), glycolysis (b), pentose phosphate pathway (c), TCA cycle (d), and lipid biosynthesis pathway (e). All experiments were performed with six replicates. Color key represents the z-score for each metabolite (normalized across all 18 samples).

During growth on cellobiose, we observed a 345-fold increase in the concentration of intracellular cellobiose, which is consistent with the upregulation of cellodextrin transporter (CDT1). Again, in alignment with the gene regulation, metabolite concentrations in central carbon metabolism exhibit mixed changes, with valine and allantoin present in higher concentrations during growth on cellobiose and others like fumarate, malate, and α-ketoglutarate are present in lower concentrations. An interesting result is decreased concentration of hexadecenoic acid, indicating a potentially lower intracellular pool of acetyl-CoA. We previously highlighted that the gene expression results also point to reduced production of acetyl-CoA. While this study did not focus on lipid production in *L. starkeyi*, it has been previously reported that *L. starkeyi* produces slightly lower lipid titers in flask cultures when grown on cellobiose, compared to glucose [135].

The metabolite concentrations during growth on xylose also mirror the gene expression results. We notice a 29-fold higher concentration of intracellular xylose. Only a few metabolites are present in lower concentrations in glycolysis and TCA cycle, indicating little difference in the regulation of central carbon metabolism between growth on xylose and glucose.

3.4 Discussion

Lipomyces starkeyi is a promising oleaginous yeast that can utilize a variety of plant-based substrates. It can grow on various hydrolysates derived from wheat straw and corn stover and exhibits tolerance to the aromatic inhibitors present in hydrolysates [116,136]. It is also an excellent lipid producer and can store triacylglycerides up to 70% of its dry cell weight. While various studies have optimized lipid accumulation using different hydrolysates, a few focus on governing metabolic pathways and physiology.

In this work, we analyze the transcriptome and metabolome of *L. starkeyi* during growth on three plant-based sugars: glucose, xylose, and cellobiose. Cell culture samples for RNA and metabolite extraction were collected during the mid-exponential phase (Fig. 3.1). The cells did not accumulate lipids, as the growth media was limited in carbon and had other nutrients in excess. We observed significant changes in the gene expression and metabolite concentration amongst the different sugars (Fig. 3.2 & 3.3) and mapped the gene regulation to relevant metabolic pathways. (Fig. 3.4-3.6).

The bulk of our analysis focuses on central carbon metabolism, which is attributed to the limited gene annotations in *L. starkeyi*. Of the 8192 genes in *L. starkeyi*, 10% are upregulated and 5% are downregulated during growth on cellobiose compared to glucose. On the other hand, growth on xylose resulted in comparatively lesser regulation: 5% of genes are upregulated, and another 1.5% downregulated. However, we observed a reverse trend from the enrichment analysis (Fig. 3.2b,c), where more pathways were enriched during growth on xylose, despite lesser differential genes, compared to cellobiose. We suspect this is because the gene expression changes in cellobiose are global. Since the gene annotation information is limited, our analysis has elaborated on the differences mainly limited to the central carbon metabolism.

We also note that some genes in the *L. starkeyi* genome were incorrectly annotated. For instance, we found 86 genes annotated as L-arabinose isomerase in the KEGG annotations. Using bioinformatics tools such as BLAST and DeepLoc, we note that all 86 genes are predicted membrane proteins, some bound to the cell membrane. A recent study found that the gene LST1_120451, marked as L-arabinose isomerase in the annotation, is a functional cellobiose transporter [119], and another gene, LST1_205437 is a co-transporter for glucose and xylose [120].

During growth on cellobiose, we saw upregulation of both cytosolic and secreted β-glucosidases, along with upregulation of a cellobiose transporter, CDT1 (Fig. 3.4a). This is in line

with cellobiose utilization in other fungi like *Neurospora crassa* [137]. We also note non-specific regulation during growth on cellobiose. Genes involved in the hydrolysis of other oligosaccharides and polysaccharides, like maltose, sucrose, maltodextrin, and β-glucan, are also upregulated. Uptake of xylose in *L. starkeyi* is also in-line with other xylose assimilating yeasts like *Scheffersomyces stipitis* [138]. Xylose is converted to xylulose through the oxidoreductase pathway (XR, XDH) and channeled into the pentose phosphate pathway primarily through xylulokinase (XKS). All three genes, *XR*, *XDH*, and *XKS*, exhibited increased expression during growth on xylose. Pentose phosphate pathway and upper glycolysis are also partially upregulated. We also observed increased expression of the genes in the glyoxylate bypass (*MLS1* and *ICL1*), phosphoenolpyruvate carboxylase (*PCK1*), and fructose bisphosphate (*FBP1*) during growth on both xylose and cellobiose, indicating the upregulation of gluconeogenesis in these growth conditions. In line with these results, we also report the increased expression of genes in the biosynthesis of the secondary metabolites like glycogen synthase (*GSY1*), glycogenin glucosyltransferase (*GLG1*), and myo-inositol dehydrogenase (*idhA*).

These results improve our understanding of glucose, xylose, and cellobiose assimilation by *L. starkeyi* NRRL Y-11557 and provide a global overview of gene expression during growth on these plant-based sugars. The associated data sets from the transcriptomics and bioinformatics analysis present potentially valuable information to help guide the metabolic engineering of *L. starkeyi* and assist in developing regulatory networks and genome-scale models.

Chapter 4: Identification of Monosaccharide Sugar Transporters[2]

4.1 Introduction

Glucose and xylose are the two most abundant sugars in lignocellulosic biomass [139]. The development of efficient and economical processes for the conversion of lignocellulosic biomass into various biofuels, chemicals and bioproducts requires microorganisms capable of utilizing both sugars if possible simultaneously [140]. Xylose metabolism, however, is not native to *Saccharomyces cerevisiae,* which has been used for the production of corn and sugarcane ethanol. A number of studies have demonstrated that *S. cerevisiae* can be engineered to efficiently utilize xylose [141–148]. However, xylose transport in these engineered strains is subject to glucose repression, which leads to sequential utilization of glucose and xylose rather than simultaneous co-utilization. Glucose repression in a xylose-fermenting engineered *S. cerevisiae* is initiated from glucose inhibition on xylose uptake by endogenous sugar transporters [149–152].

S. cerevisiae has at least 18 hexose transporters. However, dedicated xylose transporters in *S. cerevisiae* has not been reported. Xylose transport in *S. cerevisiae* is facilitated by actively expressed hexose transporters *(HXT1-7 and GAL2) as HXT8-HXT17* are either inactive (not transcribed) or cryptic [150,152,153]. Although these hexose transporters can facilitate efficient xylose utilization when it is the sole sugar, the presence of glucose completely inhibits xylose uptake due to the higher affinity of the sugar transporters toward glucose [154]. As such, glucose inhibition of xylose transport has been considered as a bottleneck preventing simultaneous co-fermentation of glucose and xylose. Several attempts have been made to bypass glucose inhibition in mixed-sugar fermentations. Ha *et al.* developed an engineered yeast strain capable of co-fermenting

[2] This chapter is adapted from the following publication: Kuanyshev, N., **Deewan, A.**, Jagtap, S. S. *et al.* Identification and analysis of sugar transporters capable of co-transporting glucose and xylose simultaneously. *Biotechnology Journal*, **16**, e2100238 (2021).

cellobiose, a dimer of glucose, and xylose, thus avoiding inhibition of xylose transport by glucose [137]. However, this strategy does not allow co-fermentation of monomeric sugars present in cellulosic hydrolysates generated by matured pretreatment and enzymatic hydrolysis processes [155,156]. Therefore, many studies have focused on identifying xylose specific transporters from xylose-fermenting yeast species, such as *Pichia stipitis* and *Candida intermedia* [157,158]. Although heterologous expression of the identified xylose transporters in a *S. cerevisiae* lacking hexose sugar transporters conferred growth on xylose, glucose inhibition on xylose transport was still observed [157,158]. In addition to bioprospecting, rational and directed-evolution approaches have led to the development of xylose transporters not inhibited by glucose [159–163]. Using rational mutagenesis, Young *et al.* reported a conserved amino-acid motif responsible for monosaccharide selectivity in sugar transporters conferring growth on xylose. Further, mutation of the conserved monosaccharide recognition motifs led to a designed transporter for xylose transport. However, the transporter could not transport glucose and xylose simultaneously, leaving the co-fermentation problem open [163]. Farwick *et al.* employed adaptive laboratory evolution of an individual sugar transporter, using a xylose-utilizing strain of *S. cerevisiae* lacking all hexose transporters and with disrupted glycolysis, to identify evolved hexose transporters insensitive to glucose repression. The authors discovered two amino-acid residues (Asn376/370 and Thr219/213) of Gal2 and Hxt7 that are essential for co-transport of glucose and xylose. However, modifying these two residues resulted in reduced rates of glucose and xylose transport [159]. Using similar approach Shin *et al.*, identified Asn366 residue mutation (same as in *Sc*Gal2/Hxt7 Asn376/370) in Hxt11 that enabled simultaneous glucose and xylose co-fermentation [162].

While the rational design approach led to promising results, we aimed to expand bioprospecting in the search of native glucose and xylose co-transporters. Oleaginous yeasts,

such as *Rhodosporidum toruloides* and *Lipomyces starkeyi* are receiving more attention as an alternative cell factory for lipid and acetyl-CoA based products given their ability to naturally consume most of the sugars including hemicellulose derived glucose and xylose [10,28]. Recently, genome sequence of *R. toruloides* and *L. starkeyi* have been reconstructed and annotated, allowing search for putative xylose transporters [41,117]. According to our xylose transporter search criteria based on conserved motif G[G/F]XXXG [163] and Thr213 and Asn370 residues [159], both species contained 8 putative xylose transporters.

In contrast to yeast transporters, the mechanism of xylose transport by SWEETs has not been studied so far. SWEETs are newly discovered family of transporters with distinct 7 transmembrane (TM) structure that plays a key role in plant development and sugar translocation within the plant phloem [164]. SWEETs are comprised by 7 TM domains, where the N-terminal three helixes shares sequence similarity to C-terminal three helixes, connected by non-conserved fourth domain [165–168]. Previous studies on *Arabidopsis thaliana* SWEETs demonstrated functional expression of the transporters in yeast, conferring growth on glucose [165,167,169]. Recently, Podolsky *et al.,* identified novel fungal SWEET from anaerobic fungi (Neocallimastigomycota) which demonstrated co-consumption of glucose and xylose in *S. cerevisiae* [170].

In this study we aimed to investigate an ability of putative xylose transporters from *R. toruloides* IFO0880 and *L. starkeyi* NRRL Y-11557 and SWEET transporters from *A. thaliana* to co-ferment glucose and xylose, a desired trait for producing cellulosic biofuels by engineered *S. cerevisiae*. In the first part of the study, we expressed selected transporters in engineered *S. cerevisiae* optimized for efficient xylose fermentation lacking major hexose transporters to screen and characterize transporters that capable to co-ferment both sugars [171]. We identified that *L. starkeyi* LST1_205437 and *A. thaliana* SWEET7 have an ability to co-ferment glucose and

xylose simultaneously. To understand kinetic background behind simultaneous glucose and xylose co-fermentation, we performed kinetic study using ^{14}C labeled sugars. Kinetics studies revealed that both transporters transports xylose in the presence of glucose. Cryo-EM or/and X ray crystallography of the selected transporters have not been resolved. Hence, to explain molecular basis of this unique trait observed in the selected transporters, we employed *in silico* molecular modelling and dynamics simulation (MD). Using crystal structure of OsSWEET2b and XylE transporters as a homology template, we performed molecular simulation of glucose and xylose transport in LST1_205437 and *A. thaliana* SWEET7.

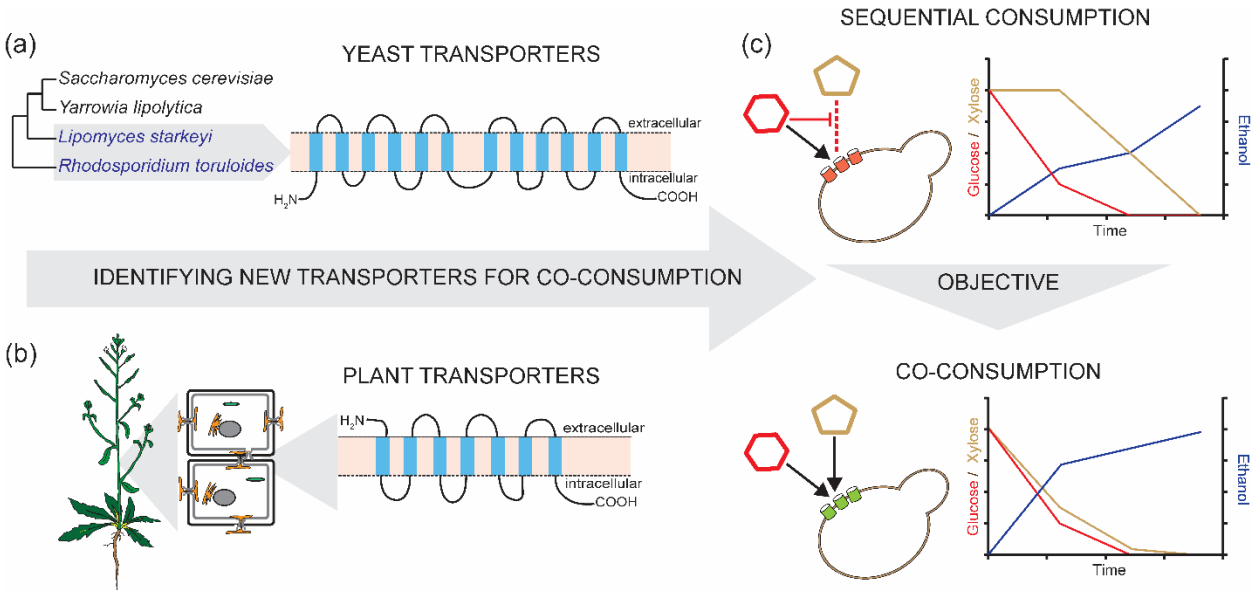

Figure 4.1. Bioprospecting strategy implemented in this study. This figure depicts the main steps applied to identify novel xylose and glucose co-transporting transporters. a Identification transporters from emerging oleaginous yeasts *Lipomyces starkeyi* and *Rhodosporidium toruloides*. b Characterization of SWEET transporters from *Arabidopsis thaliana*. c Schematic fermentation profile of a sugar mixture containing glucose and xylose by the engineered *S. cerevisiae*. Glucose presence inhibits xylose transport leading to sequential sugar utilization. Application of the discovered transporters relief glucose inhibition of xylose transport, leading to glucose and xylose co-consumption.

The study demonstrated that bioprospecting approach still can be a versatile tool to identify novel transporters with unorthodox protein motifs and residues for glucose and xylose cotransport. By combining kinetics and molecular simulation study, we were able to get insights

into a molecular basis and responsible amino acid residues enabling co-transport of glucose and xylose in LST1_205437 and *At*SWEET7 (Fig 4.1).

4.2 Materials and methods

Medium and cell growth conditions

Under non-selective conditions, all strains were grown YPD agar plates (2 % w/v agar, 1 % w/v yeast extract, 2 % peptone and 2 % glucose). A single colony from YPD agar plate was inoculated into 2 mL YPD liquid medium to obtain seed cultures. For growth study, the seed cultures were then used to inoculate 25 mL of YPD and YPX medium (10 g/L yeast extract, 20 g/L peptone, and 20 g/L xylose or glucose) in a 125 mL shake flask with a starting OD600 of 1. The cells were then grown at 30 °C and 250 rpm.

For flask fermentation, a single colony was inoculated to 5 or 25 mL YPE (1 % w/v yeast extract, 2 % peptone, 5 % ethanol) supplemented with 200 μg/ml of geneticin to obtain seed cultures. Subsequently, seed cultures were inoculated to 25 mL of YPD,YPX and YPDX medium (10 g/L yeast extract, 20 g/L peptone, and 20 g/L xylose or/and glucose) in a 125 mL shake flask with a starting OD600 of 1, 5 or 10 for flask fermentation. Flask fermentations were maintained at 30 °C and 250 rpm. $CaCO_3$ at 50g/L were added for high sugar fermentations in YPDX medium (10 g/L yeast extract, 20 g/L peptone, 70g/L glucose and 40g/L xylose).

A previously constructed xylose fermenting *S .cerevisiae* yeast (SR8) with *HXT1-7*Δ*, GAL2*Δ deletions was used for transporter screening and characterization (SR8D8) [144,171]. SR8D8 was grown in YPE medium (10 g/L yeast extract, 20 g/L peptone, and 5 g/L ethanol). The codon optimized sugar transporter genes from *L. starkeyi*, *R. toruloides* and *A. thaliana* were expressed in SR8D8 using G418 resistance dominant marker harboring plasmid for glucose and/or xylose transport characterization. SR8D8 strains transformed with plasmid containing *KanMX* marker conferring resistance to G418 (geneticin) were propagated on YPE supplemented with 200 μg/ml

of geneticin. For growth and flask fermentation experiments all media was supplemented with 200 μg/ml of geneticin for plasmid maintenance. Biomass was calculated from the OD600 measured using a Biomate 5 UV-visible spectrophotometer (Fisher, NY, USA). All growth rates were measured using a Bioscreen C plate reader system (Growth Curves USA, Piscataway, NJ, USA). A 2 μL inoculum of fully-grown culture was added into 200 μL YP containing 200 μg/ml Geneticin with varying concentrations of different sugars. A wide band filter (420–580 nm) was used to measure optical density. Bioscreen C values represent mean value from three biological replicates. In all cases, the Bioscreen C was set to maintain a temperature of 30 °C and high aeration through high continuous shaking.

Plasmid construction and transformation

All transporters were cloned into p42K-GPD1p-CYC1t plasmid harboring 2μ replication origin and *KanMX* marker conferring resistance to G418 (geneticin) antibiotic. For *At*SWEET transporters p42K-GPD1p-CYC1t plasmid were linearized with BamHI and XhoI enzymes. *AtSWEET*s were PCR amplified and digested with BamHI and XhoI. Linear p42K-GPD1p-CYC1t and AtSWEETs were ligated with T4 ligase according to manufacturer's protocol. For *R. toruloides* and *L. starkeyi* transporters p42K-GPD1p-CYC1t plasmid were linearized with BamHI and EcoRI enzymes. The trasnporters were PCR amplified and digested with BamHI and EcoRI. Both p42K-GPD1p-CYC1t and the transporters were ligated with T4 ligase according to manufacturer's protocol. All plasmid was transformed into *E. coli* DH5α for propagation and maintenance. SR8D8 yeast strain was grown on YPE medium for transformation. SR8D8 transformations were performed using LiAc method according to Gietz *et al.* [172]. Transformants were selected on YPE plate supplemented with 200 μg/ml of geneticin. *At*SWEET1 and *At*SWEET mutants were synthesized as gBlocks and cloned into p42K-GPD1p-CYC1t as

described before (Integrated DNA technologies, IA, USA). Variants of LST1_205437 mutant were synthesized from Twist Biosciences (Twist Biosciences, CA, USA) and cloned as previously described.

^{14}C labeled sugar uptake assay

SR8D8 containing the respective plasmid was grown on selective YPE medium to an OD600 of 1-1.5, harvested by centrifugation, and washed twice in ice-cold uptake buffer (100 mM potassium phosphate, pH 6.5). ^{14}C labeled sugar uptake assay was done according to Boles and Oreb [173]. Radioactivity was analyzed in a Beckman-Coulter LS6500 multi-purpose liquid scintillation counter (Beckman-Coulter, CA, USA).

Uptake was measured at sugar concentrations 0.2, 1, 5, 25, and 100 mM for glucose and 1, 5, 25, 66, 100, 200, and 500 mM for xylose. Inhibition of xylose uptake by glucose was measured at 25, 66, and 100 mM xylose with additional 25 and 100 mM unlabeled glucose. Sugar solutions contained 0.135–0.608 μCi of D-[U-^{14}C]-glucose (290-300 mCi/mmol) or D-[1-^{14}C]-xylose (55 mCi/mmol) (PerkinElmer, MA, USA). Calculation of K_m (Michaelis constant), V_{max} (maximal initial uptake velocity), and K_i (inhibitor constant for competitive inhibition) was done by nonlinear regression analysis and global curve fitting in Prism 7 (GraphPad Software) with values of three independent measurements.

Transporter identification

Orthologs of known sugar transporters were identified in *R. toruloides* and *L. starkeyi* using BlastP [174]. Glucose transporters from *S. cerevisiae* (Hxt7, Hxt2, Hxt1, Hxt3) [153,175] and xylose transporters from *P. stipitis* (Xut5, Xut2, Rgt2, Xut3) [176] were used as query sequences for blast search. Search results were filtered by e-value and gene regulation. MEGA X 10.0.1 tool [177] was

used to perform ClustalW alignment for the filtered putative sugar transporters and identify conserved structural domains and amino acid residues. The alignment results were edited using the Jalview 2.8 tool [178] for enhanced visual presentation.

4.3 Results

Identification of putative xylose transporters in *Rhodosporidium toruloides* and *Lipomyces starkeyi*

We used knowledge of existing yeast sugar transporters to identify sugar transporters in *R. toruloides* and *L. starkeyi,* which have not been searched for sugar transporters. We found multiple orthologs to HXT transporters from *S. cerevisiae* and XUT transporters from *P. stipitis*. We filtered the transporters with 12 TM domains and conserved sequence motifs (Fig. 4.2a) [157]. Recent studies have shown the involvement of the conserved motif G[G/F]XXXG [163], and Thr213 and Asn370 residues [159] in Hxt7 towards xylose specificity. As such, we used these conserved motifs and residues to refine glucose and xylose specific transporters in *R. toruloides* and *L. starkeyi*. For *L. starkeyi*, LST1_106361 and LST1_205437 were identified as putative glucose transporters and LST1_76 was identified as a putative xylose transporter. For *R. toruloides*, RTO4_11075 and RTO4_13042 were identified as putative glucose transporters, and RTO4_13731 and RTO4_10452 were identified as putative xylose transporters (Fig. 4.2c). The protein IDs' were picked from respective gene models at JGI mycocosm.

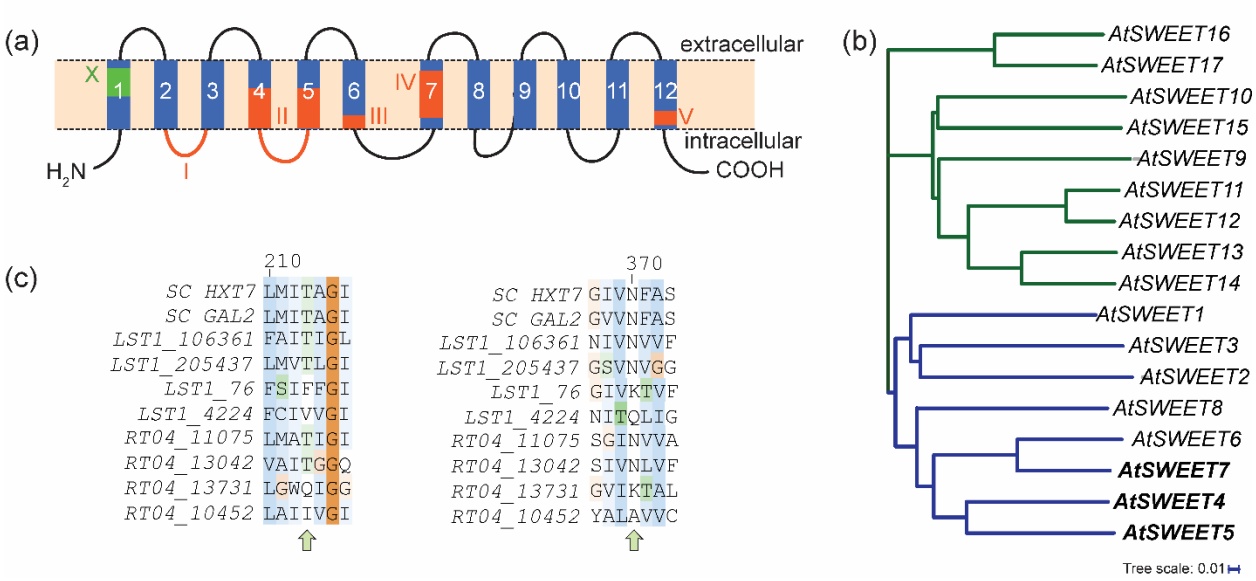

Figure 4.2. Bioinformatics analysis for transporter identification. a Most monosaccharide transporters in yeasts have 12 TM domains (represented in blue). The conserved motifs identified in yeasts transporters are marked in orange (I-V). Motif X (marked in green) has recently been identified as a key motif involved in xylose specificity. **b** A phylogenetic tree of the 17 *A. thaliana* SWEET transporters clusters the monosaccharide and disaccharide transporters independently. **c** Multiple sequence alignment of putative transporters: Thr213 and Ans370 are conserved in reported glucose transporters in yeasts.

Screening of *Arabidopsis thaliana* SWEET and oleaginous yeast transporters for glucose or xylose transport

It has been well reported that SWEETs transport different sugars, which encouraged us to examine xylose and glucose transport capabilities of 17 *At*SWEET1-17. We used an engineered *S. cerevisiae* strain (SR8D8) capable of xylose fermentation which lacks the Hxt1-7 and Gal2 transporters—rendering it unable to grow on glucose or xylose—for the examination [144,171]. We measured growth kinetics of SR8D8 transformants expressing the *A. thaliana* SWEETs and putative oleaginous yeast transporters using glucose and xylose as a sole sugar (Fig. 4.2b, Fig 4.2c, and Fig 4.3). *Sc*Gal2 expressing SR8D8 was used as a positive control. Most of the SR8D8 transformants expressing *At*SWEETs and putative oleaginous transporter were not able to grow on glucose or xylose. Only *At*SWEET4, *At*SWEET7, and LST1_205437 expressing strains exhibited robust growth on xylose and glucose (Fig. 4.3a).

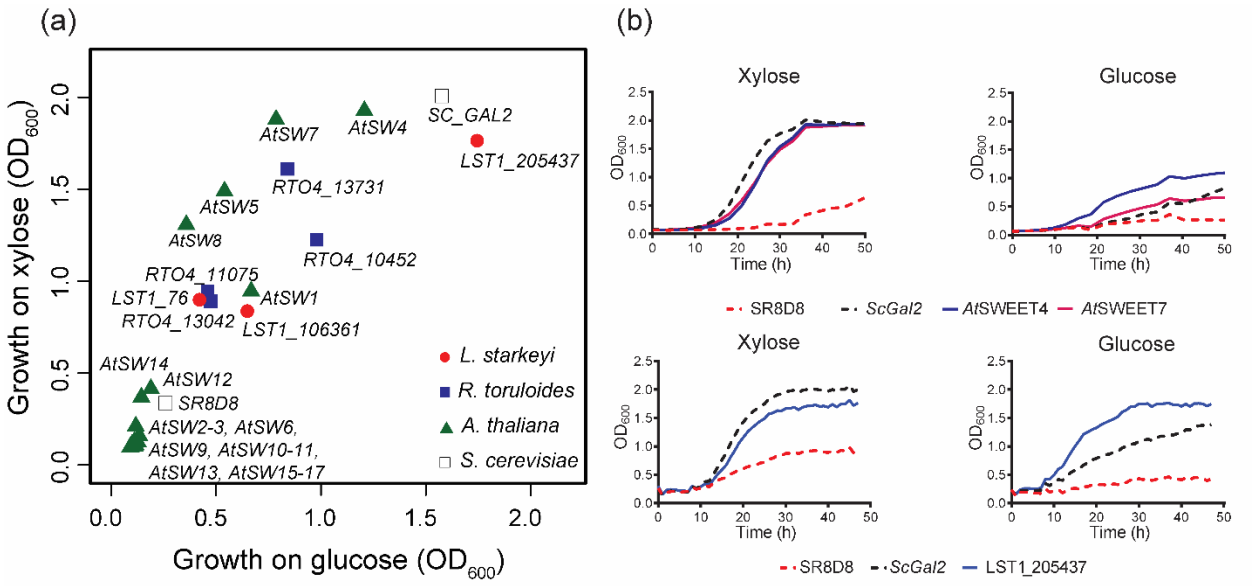

Figure 4.3. *L. starkeyi*, *R. toruloides* and *A. thaliana* SWEET transporter screening for growth on glucose or xylose. a Growth characteristics of the SR8D8 strain expressing transporters were summarized using a plot with. X axis represents the cell densities on glucose and Y axis represents the cell densities on xylose. Cell densities of the transporter-expressing strains at 40 hrs were presented. b Growth curves of the four strains with an overexpression cassette of *GAL2*, *At*SWEET4, *At*SWEET7, or a control plasmids (SRD8) on xylose and glucose. The dots and line lines are means from duplicated cultures.

A. thaliana SWEET and *L. starkeyi* LST1_205437 transporters conferred glucose and xylose cofermentation ability in engineered yeast

To test if the selected transporters can enable consumption of both sugars simultaneously upon introduction to the SR8D8 strain, we performed flask fermentations with a mixture of glucose and xylose and monitored sugar consumption over time. We used the SR8D8 expressing *GAL2* as a baseline control for determining co-consumption phenotypes, because it is known to transport both glucose and xylose in a sequential manner (Fig. 4.4a). In addition, we included *At*SWEET1 as an additional control for *At*SWEETs, because it is most studied SWEET transporter and confers growth of SR8D8 on glucose [165,179,180]. Both *At*SWEET4 and *At*SWEET7 showed simultaneous co-utilization of glucose and xylose with different rates within 24 hours. While *At*SWEET1 showed a complete preference for glucose with negligible xylose consumption, *At*SWEET4 showed co-consumption of glucose and xylose with a faster glucose

consumption rate than that of xylose. Interestingly, *At*SWEET7 enabled simultaneous co-consumption of glucose and xylose with almost same rates of sugar consumption (Fig. 4.4c). LST1_205437 transporter from *L. starkeyi* showed co-consumption of glucose and xylose (Fig. 4.4b) but glucose consumption was faster than xylose consumption.

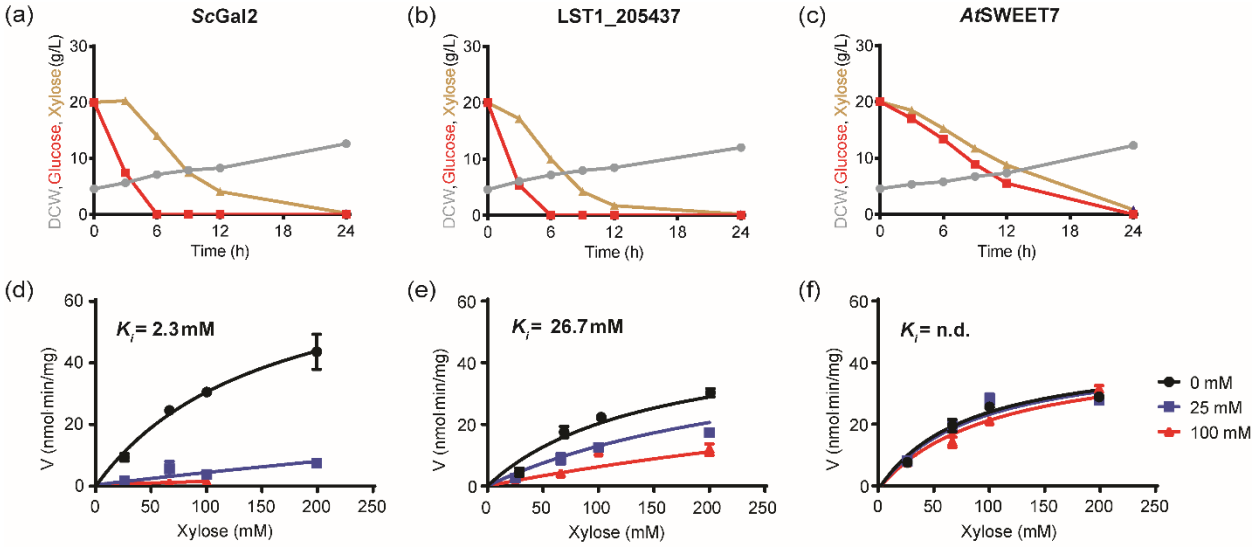

Figure 4.4. Glucose and xylose mixed sugar fermentation profile and inhibitory effect of glucose on xylose transport. 20 g/L of glucose and xylose mixed sugar fermentation by SR8D8 expressing *Sc*Gal2 (sequential fermentation) (a), LST1_205437 (partial cofermentation) (b), *At*SWEET7 (true co-fermentation) (c). Symbols: glucose (square), xylose (triangle up), DCW (circle). Inhibitory effect of 0 mM, 25 mM and 100 mM glucose on xylose transport in SR8D8 expressing *Sc*Gal2 (d), LST1_205437 (e) and *At*SWEET7 (f). Global curve fitting for Michaelis–Menten kinetics with competitive inhibition was applied to data of three independent measurements at each concentration.

In further experiments, we chose *At*SWEET1 as a sole glucose transporter, *At*SWEET7 as a glucose and xylose co-transporter, and LST1_205437 as a semi glucose and xylose co-transporter. *At*SWEET7 transports both sugars simultaneously, but suffer from slow transport capacity. While LST1_205437 performs partial co-consumption, it has an efficient transport capacity for both glucose and xylose. The difference could be attributed to the structure and function of the transporters within the isolated organism.

Next, we evaluated fermentation performances of the SR8D8 transformants expressing *At*SWEET1, *At*SWEET7 and LST1_205437 under glucose or xylose conditions. As expected, *At*SWEET7 and LST1_205437 transporters enabled glucose and xylose fermentation, depleting all provided sugars. In contrast, *At*SWEET1 enabled robust glucose fermentation but inefficient xylose fermentation with only 5 g/L of xylose consumption within 50 h.

Kinetic and molecular properties of *A. thaliana* SWEET7 and *L. starkeyi* LST1_205437

To understand kinetic and molecular basis of *At*SWEET7 and LST1_205437 glucose and xylose co-transport phenotypes, we performed radiolabeled sugar transport kinetics experiments, and *in silico* molecular modeling simulations with *Sc*Gal2 and *At*SWEET1 served as representative controls. *Sc*Gal2 was confirmed to be a high affinity glucose transporter (K_M = 1.613 mM, V_{max} = 38.33 nmol/min-mg), with low affinity toward xylose (K_M = 320.5 mM) (Table 1). Glucose transport kinetics of LST1_205437 was inferior to the *Sc*Gal2 transporter (K_M = 4.975 mM, V_{max} = 46.89 nmol/min-mg), whereas xylose kinetics was superior (K_M = 145.3 mM, V_{max} = 76.8 nmol/min-mg) (Table 1). These transport kinetic differences were not noticeable during sole sugar fermentation, unlike mixed sugar fermentation.

We then compared transport kinetic properties of *At*SWEET1 and *At*SWEET7. The results showed that *At*SWEET1 transports glucose more efficiently as compared to *At*SWEET7, with very poor xylose transport kinetics. These kinetics results of *At*SWEET1 and *At*SWEET7 are consistent with the fermentation results by the SR8D8 strains expressing *At*SWEET1 and *At*SWEET7.

Individual sugar uptake kinetics results of LST1_205437 supported the partial glucose and xylose co-consumption phenotype. However, the engineered yeast expressing AtSWEET7 showed apparent co-consumption of glucose and xylose, while kinetics results indicated

discrepancies in K_M (K_M=75mM for glucose and K_M=308mM) (Table 1). These results prompted us to directly investigate the xylose transport rates by *Sc*Gal2, LST1_205437 and *At*SWEET7 in the presence of glucose. We performed xylose uptake assay with 25 mM or 100 mM glucose, similar conditions that were used in previous study [159]. As shown in Fig. 4.4d, xylose transport by *Sc*Gal2 was completely inhibited in the presence of glucose (K_i = 2.3 mM). This kinetic behavior of *Sc*Gal2 is consistent with the mixed sugar fermentation result (Fig. 4.4a). Interestingly, xylose transport by LST1_205437 was less inhibited by glucose than those by *Sc*Gal2 (K_i = 26.7 vs 2.3 mM) (Fig. 4.4e). As a result, the LST1_205437 expressing strain showed a partial co-consumption of glucose and xylose (Fig. 4.4b). Remarkably, *At*SWEET7 showed no inhibition of xylose transport by glucose (Fig. 4.4f). Next, we performed a mixed sugar fermentation experiment under industrially-relevant sugar concentrations of 7 % glucose and 4 % xylose to validate co-fermentation of AtSWEET7 and LST1_205437. As expected the *Sc*Gal2 expressing strain exhibited a sequential utilization of glucose and xylose (Fig. 4.5a). The sugar utilization profile of the LST1_20437 expressing strain was consistent with the kinetics data, showing partial xylose and glucose co-consumption (Fig. 4.5b). The AtSWEET7 expressing strain showed co-consumption of glucose and xylose even at higher glucose concentrations, further supporting that AtSWEET7 is indeed glucose and xylose co-transporter which is insensitive to glucose inhibition even under high glucose concentrations (Fig. 4.5c).

Alteration of Asn365 amino acid residue in *L. starkeyi* LST1_205437 changes sugar preference

Asn370/376 residues in *S. cerevisiae* hexose transporters Gal2 and Hxt7 play a critical role in glucose and xylose co-transport [159]. Replacing the Asn370/376 residue in Gal2 and Hxt7 with either hydrophobic or hydrophilic amino acids led to alleviation of glucose inhibition on xylose transport [159]. Interestingly, LST1_205437 transporter retains Asn365 (equivalent to Asn370 in

Gal2) residue and show partial inhibition of xylose uptake by glucose (Fig. 4.4b and 4.4e). We sought to test if alteration of Asn365 residue in LST1_205437 to phenylalanine, serine or valine would further alleviate glucose inhibition on xylose transport, allowing complete co-fermentation of glucose and xylose. We found that Asn365Phe, Asn365Ser, and Asn365Val mutations in LST1_205437 resulted in similar phenotypic changes as it was reported by Farwick e*t al*. Particularly, Asn365Phe mutation abolished glucose transport while retaining xylose, Asn365Ser and Asn365Val showed co-fermentation phenotypes. Altogether Asn365 residue mutation functions not only in *S. cerevisiae* transporters but also in *L. starkeyi* LST1_205437, supporting the universal importance of Asn370/376 residue in closely related yeast hexose transporters.

4.4 Discussion

The wealth of sequencing information and recently discovered SWEET family sugar transporters are still unexplored by bioprospecting for tackling glucose and xylose co-transport problem. In this study, we undertook a bioprospecting approach to identify glucose and xylose co-transporting transporters from unexplored oleaginous yeasts and plant (Fig. 4.1). We identified 8 putative xylose transporters in *R. toruloides* and *L. starkeyi* (Fig. 4.2c). However, experimental validation of the putative transporters using a xylose-fermenting *S. cerevisiae* lacking major hexose transporters (SR8D8) showed that only *L. starkeyi* LST1_205437 can enable robust growth on either glucose or xylose (Fig. 4.3a). Interestingly, LST1_205437 retained conserved Thr213 and Asn370 residues, and demonstrated a partial cofermentation of glucose and xylose (Fig. 4.4b). Furthermore, the glucose inhibition kinetics by LST_205437 showed less glucose inhibition on xylose transport whereas *Sc*Gal2 exhibited severe glucose inhibition on xylose transport even under a low glucose concentration (25mM) (Fig. 4.4d,e). This

observation provides evidence that other than Thr213 and Asn370 residues might be involved in the partial cofermentation phenotype. The fermentation experiments also support our prediction and mutation of Ala335Asn decreases the xylose uptake in LST1_205437.

Most studies related to xylose transporters focused on MFS (**M**ajor **F**acilitator **S**uperfamily) type transporters with 12 TM domains, and other families of sugar transporters have been overlooked. Here, we expanded bioprospecting approach toward SWEET family transporters. *A. thaliana* has 17 SWEET transporters that can transport either monosaccharides or disaccharides across a membrane via concentration gradients (Fig. 4.2b) [181]. According to Han *et. al. A. thaliana* SWEETs can be divided into two distinct groups based on conserved residues dictating sugar preference to monosaccharide or disaccharide. However, the authors discovered that this division could not reflect sugar specificity for all *At*SWEETs. In particular, Han *et al.* showed that *At*SWEET13 have both glucose and sucrose transport activities [166]. Therefore, in this study, we screened all 17 *At*SWEETs to identify xylose and glucose transporter. Interestingly, 17 *At*SWEETs share sequence similarity and yet showed very different sugar uptake phenotypes on glucose or xylose. We confirmed *At*SWEET1 to be a glucose transporter with almost no xylose transport capacity, whereas *At*SWEET4 and *At*SWEET7 showed both glucose and xylose transport capacities (Fig. 4.4c). Moreover, among screened transporters, *At*SWEET7 exhibited complete co-fermentation phenotype. The kinetic analysis of *At*SWEET7 revealed no glucose inhibition of xylose transport, though the glucose and xylose transport kinetic properties were poorer than *Sc*Gal2 and LST_205437 (Fig. 4.4f). Moreover, AtSWEET7 exhibited complete co-fermentation of glucose and xylose even at high residual glucose concentrations, suggesting the transporter is completely insensitive to glucose inhibition (Fig. 4.5c). Recently, Podolsky *et al.*, demonstrated utility of fungal SWEET transporters to tackle glucose and xylose cotransport problem. The authors demonstrated that the wild-type NcSWEET1 and the best performing

chimera derived from it allowed co-transport of glucose and xylose. However, in their experimental setup *S. cerevisiae* expressing wild type and the chimera transporter co-consumed only 20 g/L of sugars within 120 hours [170]. Similar results were achieved in engineered Asn366Thr Hxt11 transporter, which belongs to MFS family, engineering of native glucose and xylose co-transporter with more simpler molecular structure than MFS might be advantageous for transporter engineering [162].

In summary, this work demonstrates how bioprospecting can identify unique transporters for industrial applications. Availability of vast amounts of sequencing information, allowed us to identify and characterize yeast transporter LST_205437 that has partial glucose and xylose co-consumption capacity. We found that LST_205437 has non conserved amino acid residue responsible for the phenotype. We characterized newly discovered SWEET transporters, which are structurally different from its yeast counterparts. Altogether, information gathered in this study will increase the understanding of yeast hexose transporters and SWEET transporters, providing valuable information for industrial biotechnology and fundamental biology.

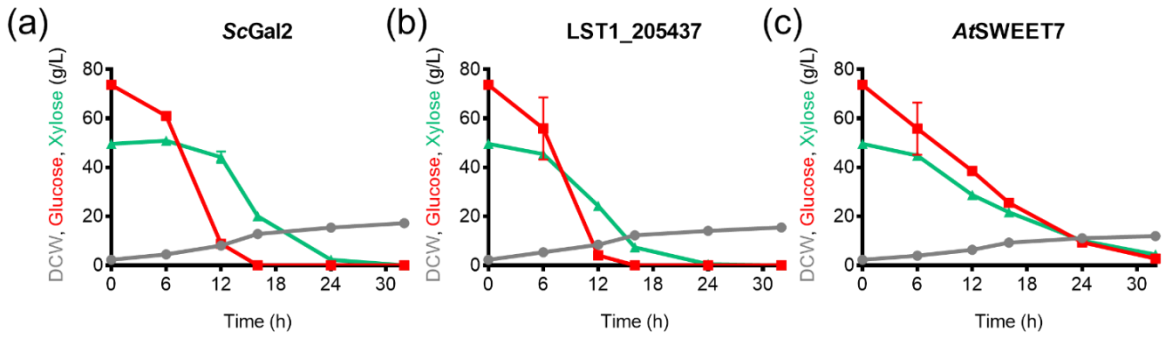

Figure 4.5. Glucose and xylose mixed sugar fermentation profile using industrially relevant sugar concentrations. 70 g/L of glucose and 40g/L xylose mixed sugar fermentation by SR8D8 expressing *Sc*Gal2 (a), LST1_205437 (b), *At*SWEET7 (c). Symbols: glucose (square), xylose (triangle up), DCW (circle). The values are the means of two independent experiments, and the error bars indicate the standard errors.

Chapter 5: Copy Number Variation in *Yarrowia Lipolytica* Clones

5.1 Introduction

Large amounts of fossil-derived fuels and chemicals are consumed worldwide every year, causing problems such as global warming and limited supply sustainability [182–185]. Therefore, renewable energy sources are of great interest and have been widely studied. Using yeasts as cell factories to produce value-added products offers a promising and sustainable solution to issues caused by petroleum-based products [186]. Oleaginous yeasts can produce elevated amounts of lipids compared to traditional yeasts, offering a feasible platform for industrial production of fatty-acid-derived products [94,187–189]. They can accumulate high amounts of lipids up to 50% of their dry weight [94,190]. *Y. lipolytica* is one of the most studied oleaginous yeasts due to advances in molecular genetic engineering tools [191].

Long chain fatty acids are produced naturally by oleaginous yeasts as energy and carbon storage molecules [189], or as precursors for building cell membranes [192], whereas medium chain fatty acids (C6-C12) are rarely produced metabolites in most organisms [189]. However, compared to long chain fatty acids, medium chain fatty acids are considered more desirable substrates for many industrial and commercial applications because the associated alkyl chain lengths are similar to those found in typical petroleum-derived products [192,193]. Therefore, engineering oleaginous yeast to produce medium chain fatty acids leads to production of renewable precursor molecules more suitable for petroleum-derived products. Medium chain fatty acids have additional uses as herbicides, precursors to lubricants and polymer additives with a much higher selling price [193,194]. Engineering oleaginous yeasts to produce medium chain fatty acids leads to the production of renewable substrates to replace petroleum-derived products [189].

In *Y. lipolytica*, fatty acid synthesis takes place in the multienzyme fungal type I fatty acid synthase (FAS) [195,196]. Initiation of fatty acid synthesis occurs when malonyl-CoA is formed from condensation reaction of acetyl-CoA with carbon dioxide by acetyl-CoA carboxylase [189]. Then transacylase activates malonyl-CoA for reaction with actyl-CoA to produce malonyl-acyl carrier protein (ACP). In bacteria, plant, and animal cells, thioesterase is a key determinant of fatty acid chain length, terminating the elongation cycle by hydrolyzing of the acyl esterified on ACP, whereas in yeast, no thioesterase is present and instead the elongation cycle is terminated via malonyl palmitoyl transferase (MPT) that transfers the acyl-ACP to a CoA [196]. In recent reports, coupling the fungal FAS system with a thioesterase enables the production of medium chain fatty acids in *Y. lipolytica* [189,197] and short chain fatty acids in *Saccharomyces cerevisiae* [195]. A heterologous thioesterase was able to associate with the native FAS complex in *Y. lipolytica* to prematurely terminate fatty acid elongation cycle [192,198].

Genome integration of heterologous DNA is preferable for stable expression and genetic modification [199]. The most frequently used method for genome integration is homologous recombination-based genome integration into specific chromosomal sites. However, very low homologous recombination efficiency in most nonconventional yeast including *Y. lipolytica* limited the application for targeted integration [199,200]. CRISPR-Cas9-based genome editing tools were developed for targeted and markerless gene deletion and integration in *Y. lipolytica* [201–203]. Despite the previous efforts to improve the homologous recombination efficiency, constructing and optimizing large and multiple biochemical pathways still remains difficult in *Y. lipolytica*.

In previous works, *Y. lipolytica* was engineered to produce medium chain fatty acids from glucose by random integration of a heterologous thioesterase [189,204]. However, homologous-independent integration leads to uncertainty in the chromosomal locations and gene copy number of the inserted heterologous gene fragments, therefore resulting in varying gene expression and

enzyme activity [199]. We discovered large variation in medium chain fatty acid accumulation across the clones constructed. We identified integration sites and gene copy number by both genome sequencing and Genome Walking kit. Interestingly, genome sequencing analysis enables the estimation of copy numbers and the fraction of insertions of integrated fragments easily. Then, we investigated more closely on how integration sites and copy number could affect thioesterase expression and medium chain fatty acids production. Expression up to 7-fold difference was seen in thioesterase expression and high variation was seen in the fatty acid accumulation profile, using quantification by RT-qPCR and GC-MS.

5.2 Materials and methods

Strains and growth conditions

All strains are derived from *Y. lipolytica* W29 (ATCC 20460). For targeted integration, *Y. lipolytica* PO1f, a leucine and uracil auxotroph variant of W29 strain was used as the host. YPD medium (20 g/L glucose, 20 g/L peptone, and 10 g/L yeast extract) was used for routine growth. Low-nitrogen medium (50 g/L glucose, 1.7 g/L yeast nitrogen base without amino acids or ammonia sulfate, 0.15 g/L yeast extract, and 1g/L ammonium sulfate) was used to induce lipid production. The cells were first grown overnight for 16 h in YPD medium at 30 C and 200 rpm. The cells were then sub-cultured in low-nitrogen medium to a starting OD600 of 0.2. The samples were incubated at 30 ºC and 200 rpm in 250 mL baffled flasks for 96 h prior to analysis. All experiments were performed in triplicate.

Plasmid and strain construction

Table 1 lists all strains and plasmids used in this study. The expression vector was constructed previously in our lab [189,204]. Heterologous fatty acyl-ACP thioesterase from *Cuphea palustris* was

randomly integrated in *Y. lipolytica* to accumulate medium-chain fatty acids. All plasmids were grown and maintained in LB medium (10 g/L tryptone, 5 g/L yeast extract & 10 g/L sodium chloride) supplemented with 100 ug/ml ampicillin in the E. coli strain NEB5α (New England Biolabs, Cat. # C2987). Two Y. lipolytica strains were used, W29 (ATCC 20460) and PO1f-Iku70 which is PO1f (ATCC MYA2613) with an inactivated ku70 gene. YPD media (20 g/L glucose, 20 g/L peptone & 10 g/L yeast extract was used for routine growth and when necessary supplemented with 150 ug/ml hygromycin or 5-flouroorotic acid (5FOA) at 1 mg/ml. YSC media.

In addition, the thioesterase cassette was specifically inserted at A08 site in PO1f chromosome, using CRISPR-Cas9 system [202]. Two gBlocks were ordered from Integrated DNA Technologies (IDT). Ptub-HPH contained a Klebsiella pneumoniae hphB gene codon optimized for use in Y. lipolytica (Genbank Accession No: KU561939) (Rutter et al., 2015) under the control of the native Y. lipolytica TUB1 promotor. Ptef-CpaFAT-Tcyc contained the Cuphea plaustris fatty acyl-ACP thioesterase (FAT) optimized for use in Y. lipolytical (Genbank Accession No: KR180392) (Rutter & Rao, 2016) flanked by a native Y. lipolytica TEF intron promotor and native CYC1 terminator. Both gBlocks contained 20 base pairs of homology to their respective insertion sites on both the 5' and 3' ends for ligation using Gibson.

pHPH (MPS132) was built by digesting pUC19 with SmaI and fusing it to the pTUB-HPH gBlock using Gibson.

pHPH was then digested with KpnI and fused to the Ptef-CpaFAT-Tcyc gBlock using Gibson resulting in the plasmid pHPH-FAT (MPS135).

Gibson assembly was performed using the NEBuilder HiFI DNA Assembly Kit from New England Biolabs (Cat. # E5520) according to the instructions provided with the kit.

The plasmid pHR-A08-hrGFP (Schwartz, et al., 2017)) was digested with BssHII and NheI then run on a 1.0% agarose gel. The 8.1 kB pHR-A08 backbone was excised and run over a Zymoclean Gel DNA Recovery column (Zymo Research, Cat. # D4001).

The Ptef-CpaFAT-Tcyc cassette was PCRed from pHPH-FAT using Phusion High-Fidelity DNA polymerase (NEB, Cat. # M0530S) and the primers MP163 & MP164. The primers also added a 5' BssHII site and a 3' NheI site. The PCR product was run over a Zymo Clean & Concentrator column (Zymo Research Cat. # D4013. 1 ug was digested with BssHII and NheI then run over another Zymo column. Finally the pHR-A08 plasmid backbone and Ptef-CpaFAT-Tcyc PCR fragment were joined using T4 ligase (NEB, Cat. # M0202S). Integrity of the Ptef-CpaFAT-Tcyc insert was confirmed by sequencing.

Transformation

Random integration into W29**:** pHPH-FAT was linearized with SfoI then transformed directly into W29 cells using a One-step method (Chen et al., 1997). Briefly, W29 cells were streaked onto a YPD plate and grown overnight at 30^{o}. Cells were scraped from the plate, resuspended in water and counted. For each transformation being performed 5 x 10^{7} cells were removed and pelleted. The pellet was resuspended in 100 ul of transformation mix per transformation. 100 ul of cells were transferred to PCR tubes and 100 ng linear plasmid DNA was added to each. Transformations were heated at 39^{o} for 1 hour in a thermocycler. Cells were centrifuged, resuspended into 100 ul of water then plated onto YPD + 150 ug/ml hygromycin and grown at 30^{o} for 2-3 days.

Targeted integration into A08 locus: One microgram each of pHR-A08-FAT and pCRISPRyl-A08 (Schwartz et al.) were cotransformed into the Y. lipolytica strain PO1f-Iku70 using the One-step method outlined above. However, post-transformation the cells were spread onto YSC -leu –ura plates and grown at 30^{o} for 2 days. Sixteen colonies were screened for

correct integration using PCR. Two of the positive colonies were inoculated into YPD + 1 mg/ml 5FOA and grown overnight at 30^0 to remove the pHR-A08-FAT and pCRISPRyl-A08 plasmids. A sample of these cultures was then spread onto YPD plates. After 3 days 8 subclones were checked for integration of FAT into the A08 locus using PCR. One of the 8 clones showed proper integration. This clone, MPS171, was used in all future experiments.

Transformation mix: 45% PEG 4000, 0.1M DTT, 0.1M Lithium acetate & 25 ug/100 ul single stranded salmon sperm DNA.

One Step reference: Chen, et al. Appl. Microbiol Biotechnol (1997) 48:232-235. One-step transformation of the dimorphic yeast Yarrowia lipolytica

Schwartz, et al. **ACS Synth. Biol.** (2017) 6, 3, 402–409. Standardized Markerless Gene Integration for Pathway Engineering in *Yarrowia lipolytica*

Lipid extraction and quantification

Total lipid quantification was performed using a sulfo-phospho-vanillin (SPV) colorimetric assay. Briefly, 50 µl cell suspension was washed twice, mixed with 1 mL concentrated sulfuric acid in glass Pyrex tubes and heated for 10 min at 100 ºC. After cooling the reaction for 10 min at room temperature, 2.5 mL of freshly prepared vanillin-phosphoric acid was added and reacted for 15 min at 37 ºC. Tubes were cooled at room temperature for 10 min and absorbance is measured at 530 nm. Corn oil (50 – 250 µg) was used to generate a reference curve, from which the lipid concentration was determined.

Fatty acids profiles were measured using GC-MS. Neutral and charged lipids were converted into their respective methyl esters by saponification for the analysis by gas chromatography/mass spectrometry (GC/MS). Lyophilized samples were resuspended in 2 mL of a 20:1

methanol/acetyl chloride mixture and 2 mL of hexane, with 5 μL of 25 mg/mL tridecanoic acid dissolved in a 3:2 methanol/benzene mixture added as an internal standard. The samples were boiled for 30 minutes in sealed glass tubes. Next 1 mL of water was added to allow phase separation, with upper organic phase containing total saponified cell lipid collected.

GC/MS was used to determine the length and abundance of the lipid species. Lipid samples were analyzed using a DB-5ms capillary column (Agilent) on a Shimadzu QP2010 GC-MS. 1 μL samples were injected at a 10:1 split ratio using hexane as the solvent. Helium carrier gas was used at a pressure of 121.7 kPa with flow rate of 1.0 mL/min. The injection port temperature was set to 250 ºC. Column temperature increased from 30 ºC to 250 ºC at a rate of 10 ºC/min. The effluent from the GC entered the ionization chamber at 250 ºC and measured at a full scan between 15 and 250 amu. Using standards as a reference, concentrations of lipid species were determined.

FAT cassette Identification by whole genome sequencing

The genomic libraries were prepared with the Hyper Library construction kit from Kapa Biosystems (Roche). The libraries were sequenced on a HiSeq 4000 (Illumina), with 2×100 paired-end reads targeting a genome coverage of 200x per sample. Fastq files were generated and demultiplexed with the bcl2fastq v2.20 Conversion Software (Illumina). Fasta files were uploaded to NCBI (BioprojectID: PRJNA732941, NCBI Accession number: SRR14663210 to SRR14663213).

Adapters and low-quality reads were trimmed using Trimmomatic. Trimmed reads were mapped onto Yarrowia lipolytica W29 reference genome (RefSeq assembly accession: GCA_001761485.1) using bwa. Samtools was used for manipulation of alignment files and read depth analysis. IGV was used for alignment visualization. Read counts and integration location

were analyzed using R 4.0.5 (R Core Team, 2021), the GenomicRanges (v 1.42.0) and Rsamtools (v2.6.0). Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. https://bioconductor.org/packages/Rsamtools) packages. The ggplot2 (v3.3.3), ggpubr (v0.4.0).packages were used for plotting. *Y. lipolytica* genome models and functional annotations were obtained from NCBI (RefSeq assembly accession: GCA_001761485.1).

FAT cassette identification by Genome Walking kit

The sites of the integrated cassette were characterized using a Genome Walking Kit (TaKaRa). The cells were cultivated in YPD medium for 2 days at 30 ºC and the genomic DNA was extracted using Dr. GenTLE High Recovery DNA extraction kit (TaKaRa Bio). The genomic DNA was digested to create four DNA libraries with four different digestive enzymes (EcoRV, DraI, PvuII and Ssp I). Then both ends of the blunt-end DNA fragments were appended with GenomeWalker Adaptor of known sequence by ligation. Long-distance PCR was used to amplify the fragments with one primer in the adaptor and the other in the integrated cassette. The amplified PCR fragments contains sequence from both the cassette and the upstream/downstream region. The upstream/downstream sequence is blasted against the wild type W29 genome and the loci for insertion is therefore determined. A PCR of the genomic DNA is performed to verify loci and size of the insertion.

FAT expression quantification by RT-qPCR

RT-qPCR is used to determine the expression of the acyl-ACP thioesterase for each clone. Cells were first cultivated in low-nitrogen medium with C/N ratio of 100 after 24 hours of growth. We kept the total OD the same for RNA extraction so that total number of cells is approximately constant for each clone. The RNA was extracted using TRIzol reagent following

the provided protocol (Invitrogen). Immediately after RNA extraction, cDNA was synthesized from RNA using M-MLV reverse transcriptase (NEB). Then the cDNA product was quantified in qPCR-reaction using Luna Universal qPCR Master Mix with at least three replicate samples (NEB). Actin (ACT1) expression was also determined as reference for relative quantification. Acyl-ACP thioesterase expression was quantified against ACT1 expression for each cloned strain. In addition to random integrated clones, acyl-ACP thioesterase expression for PO1f strain with targeted integration at A08 site was also determined using the same approach described above.

5.3 Results

Variations in total lipid and medium chain fatty acids accumulation

Most microorganisms predominately produce long chain fatty acids with acyl chains of 16 or 18 carbons in length. In an attempt to produce shorter alkyl chains, we expressed a heterologous acyl-ACP thioesterase enzyme with specificity toward acyl-ACP chains eight carbons in length. We evaluated their ability to produce C10 fatty acids among 20 clones with acyl-ACP thioesterase randomly integrated (Fig. 5.1). We found that the clones varied significantly in the amount of C10 fatty acids produced. The highest one was Clone 19 with C10 fatty acid titer up to 300 mg/L, whereas several clones produced little or no C10 fatty acid. As a reference, the *Y. lipolytica* W29 wild type strain showed no production of C10 fatty acid production. Because of random integration, acyl-ACP thioesterase was integrated in different chromosomal sites and with variable copies in different clones, leading to different expression of the enzymes.

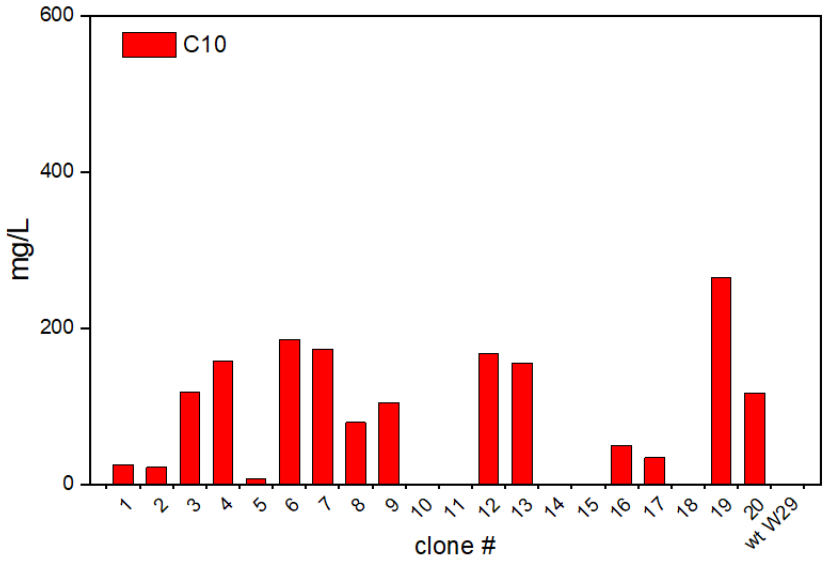

Figure 5.1. Production of decanoic (C10) fatty acids in engineered *Y. lipolytica* as determined using GC-MS.

Among the twenty clones we screened, we picked four that would reflect variance of the population, with low C10 producer Clone #1, medium C10 producers Clone #7 and #20, and high C10 producer Clone #19. We next measured the total lipid titers during growth under lipid accumulation conditions (Fig. 5.2). Total lipid titers from engineered strains expressing heterologous acyl-ACP thioesterase were comparable to those of the wild type strain. One exception is Clone #1 which produced a slightly lower total lipid titer compared to others.

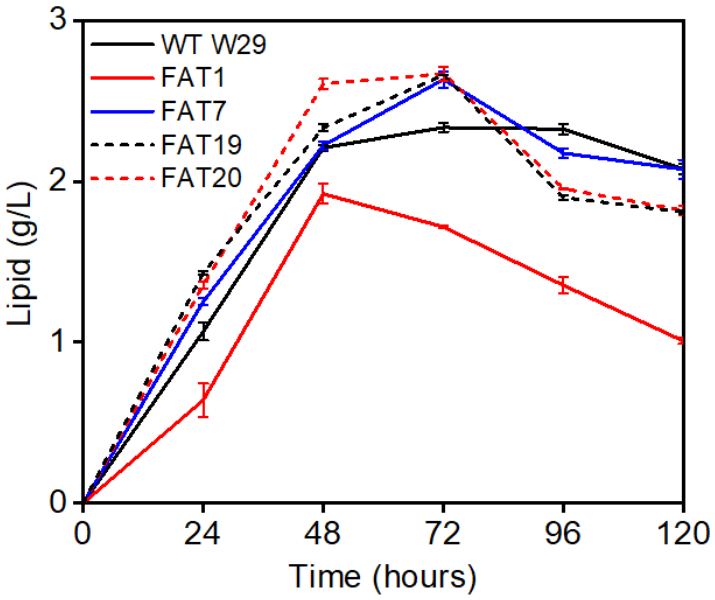

Figure 5.2. Production of total lipid in engineered *Y. lipolytica*.

The four clones were tested for their ability to produce medium-chain fatty acids (Fig. 5.3). We found that Clone #7, Clone #19 and Clone #20 were able to produce significant amount of medium-chain fatty acids (octanoic C8 and decanoic C10 fatty acids). Interestingly, Clone #20 showed a higher C8 than C10 fatty acid accumulation. In comparison, Clone #1 was not capable of producing medium-chain fatty acids. The lipid length profile for Clone #1 resembled that for the wild type strain, producing predominately fatty acids with 16, 18 and 20 carbons in length. The accumulation of 16-carbon lipids showed little variation among different strains, but 18-carbon and 20-carbon lipids showed significant reduction in strains that produced medium-chain fatty acids. Clone #20, which produced the largest amount of C8 lipids, showed the greatest reduction in production of 20-carbon lipids.

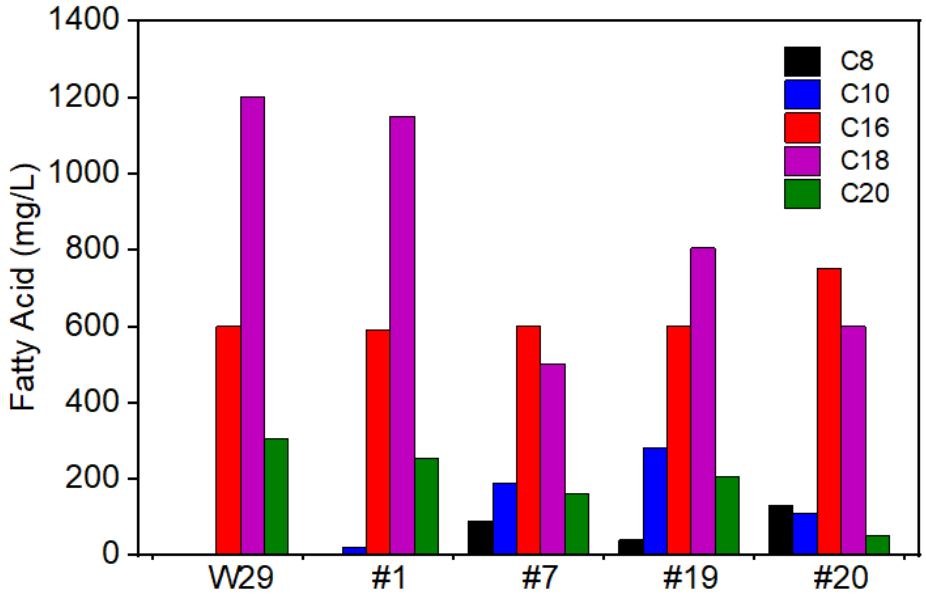

Figure 5.3. Lipid length profiles in *Y. lipolytica* strains expressing acyl-ACP thioesterase enzymes.

Detection of gene location and copy number of homology-intendent integration

From the whole genome sequencing, we were able to identify the copy number and integration locations of the plasmid. Blast was used to identify the parts of the plasmid that overlap with the *Y. lipolytica* W29 reference genome, namely pTub1 and TEF intron. The two regions were deleted from the reference genome to avoid mapping duplications and the plasmid was added as a separate chromosome before mapping. The circular plasmid was linearized at the expected cut location, location 237 in the sequence. Read depth was calculated for all locations in the plasmid using samtools. Mean read depth of chromosome 1a was used to normalize the read depth of the plasmid. Fig. 5.4 represents the read depth ratios of plasmid for different chromosomes.

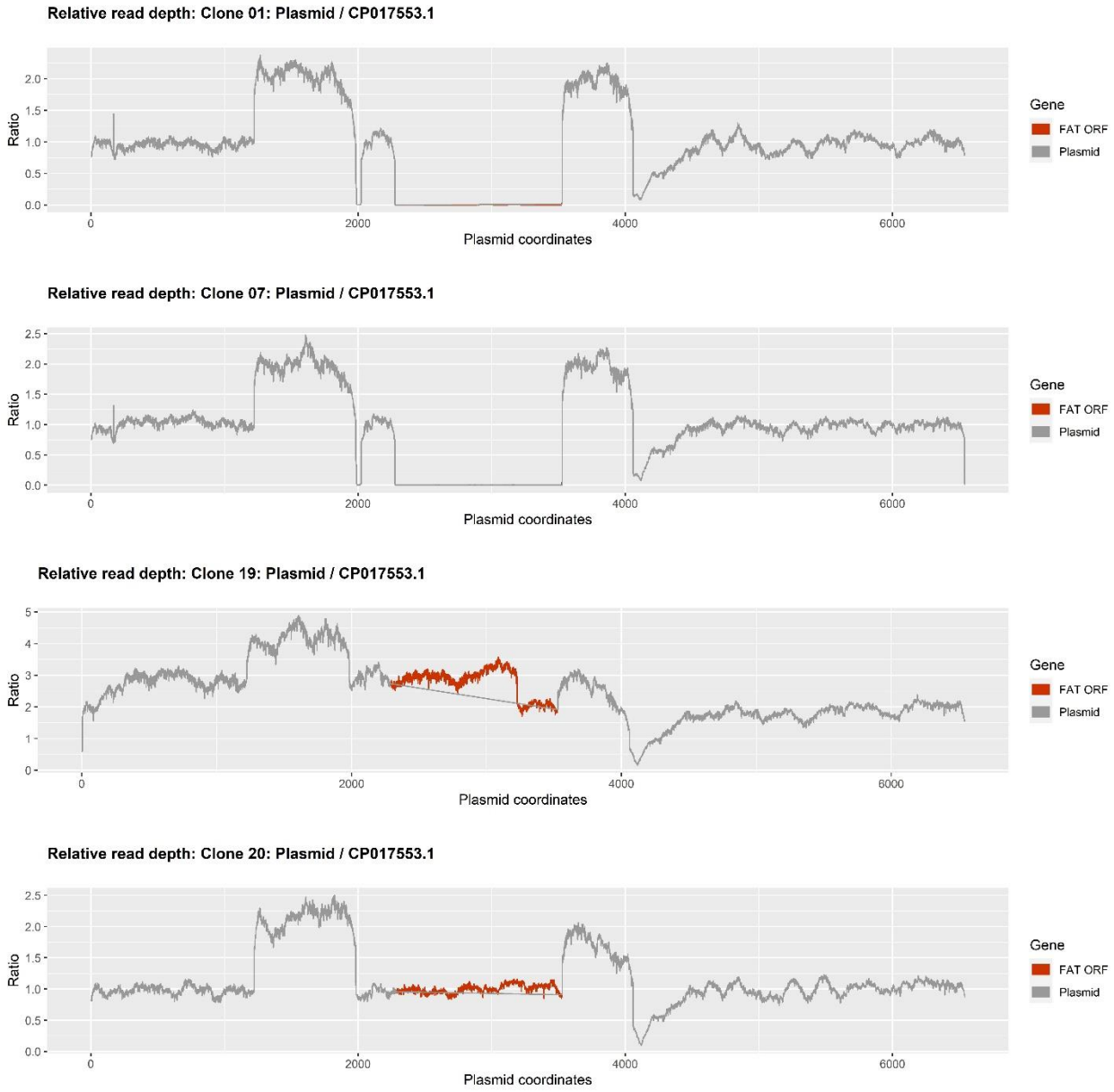

Figure 5.4. Mapping from genome sequencing.

Genome mapping indicates that clone 20 has single copy of FAT, whereas clone 19 has multiple copies – 2 whole genes and one partial copy. Clone 1 and 7, have the plasmid integrated but no copies of FAT are present. The regions that represent pTub1 (1226 - 1982) and TEF intron (3527 - 4058) are expected to show an extra read copy as compared to the rest of the

plasmid, as the reads that would have mapped to the reference genome (deleted for the purpose of this analysis) are also mapping to the plasmid.

The alignment files were used to find the locations of plasmid integration. When one portion of an NGS read maps to one location and other portion of the same read maps to a different location of a genome, the resultant read is called a split read. For this analysis, we looked for split reads that map both to the plasmid and one of the other chromosomes. The mapping location of these reads was plotted for different chromosomes. (Fig. 5.5).

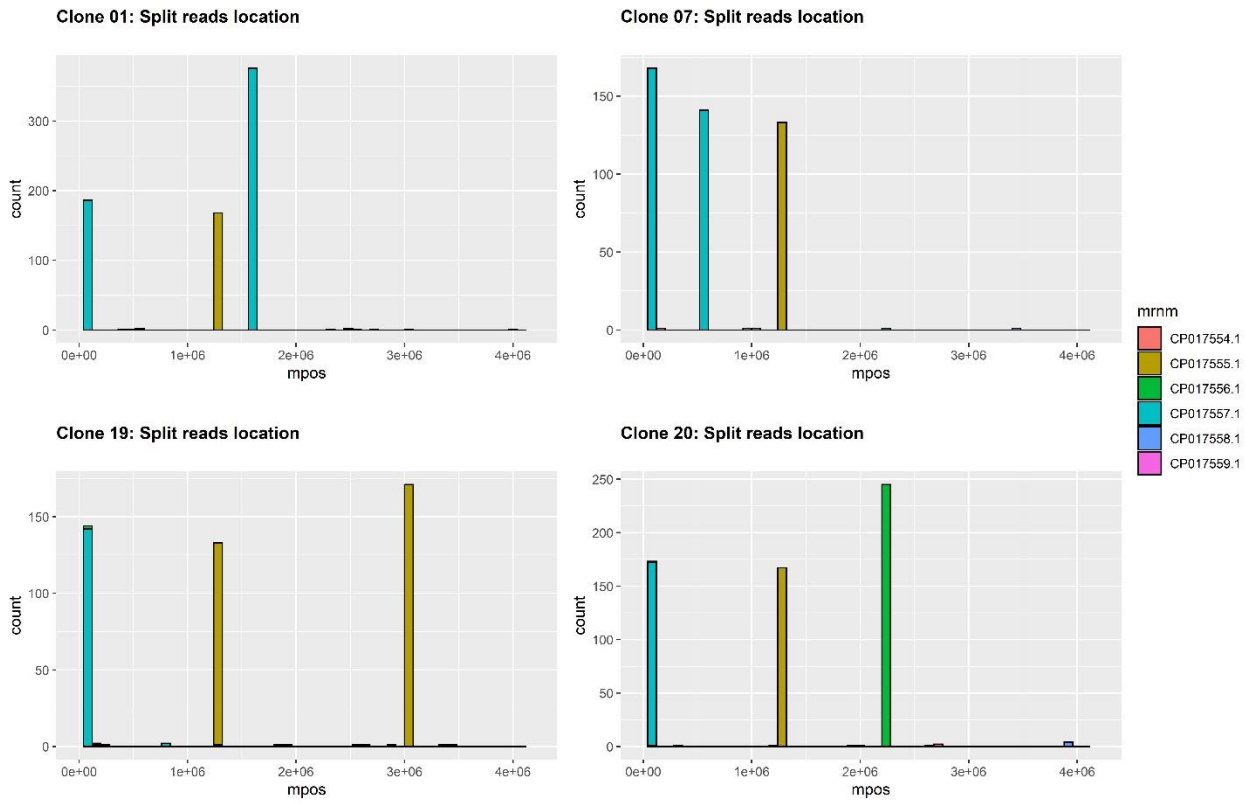

Figure 5.5: Location of plasmid integration.

All clones had plasmid fragments that mapped to chromosomes 1c (near coordinate 126400) and 1e (near coordinate 110000). These correspond to the locations of TEF intron region and ptub respectively. Fragments from both clone 1 and 7 mapped to different locations in chromosome 1e. Clone 1 mapped near YALI1_E16164p, preprotein translocase subunit Sec66.

Clone 7 mapped between YALI1_E05907p (similar to DED81, asparaginyl-tRNA synthetase) and YALI1_E05929p (Prolyl-tRNA synthetase). Additionally, clone 19 had another mapping location in chromosome 1c, before the coding region of YALI1_C30760p. Clone 20 also mapped to chr 1d, after the coding region of YALI1_D22018p. The function of these proteins is unknown according to annotations and protein blast against other yeasts in the NCBI database. We were unable to characterize, using genome sequencing, the exact fragments in which the plasmid separated and the specific locations of each fragment since the number of filtered split reads were limited (100-400).

The genome sequencing indicated that Clone #19 contained two functional gene copies and Clone #20 contained one gene copy.

We also attempted to determine the location and gene copy number in Clone #7 and Clone #19 using Genome Walking kit as an alternative verification of genome sequencing analysis. Chromosomal DNA were fragmented by digestion and adapters were added to both ends of each fragment. Using primers in the adapter and integrated gene, the chromosomal location can be therefore determined by sequencing the amplifying region. We were able to identify the location of integration for Clone #7. In Clone #19, however, the sequencing results showed mixed PCR fragments, suggesting multiple gene copies present. The mixed PCR fragments were difficult to further separate, possibly due to regions with more than 85% GC content.

Acyl-ACP Thioesterase quantification using RT-qPCR

The four selected strains showed great variation in lipid length profile, suggesting the probable cause by variation in the amount of acyl-ACP thioesterase expressed. We investigated variation in the protein expression of acyl-ACP thioesterase by homology-independent integration in the four selected clones using quantitative PCR (qPCR) in Fig. 5.6. In addition, the

wild type PO1f strain integrated with acyl-ACP thioesterase at A08 site was also quantified for its protein expression for comparison. Actin expression was used as a reference for relative quantification. In all five strains, acyl-ACP thioesterase were expressed in high amounts compared to actin expression. Acyl-ACP thioesterase expression in the five strains also showed great variation up to 7-fold. This pattern matched with the lipid length profiles. Clone #1 showed the lowest acyl-ACP thioesterase expression and produced negligible medium-chain fatty acids. Clone #20 produced medium-chain fatty acids comparable to Clone #7 and #19, but with higher acyl-ACP thioesterase expression. However, Clone #20 noticeable higher C8 fatty acids and lower C20 fatty acids, which can be caused by the high acyl-ACP thioesterase expressed.

5.4 Discussion

Expression of heterologous acyl-ACP thioesterase enabled *Y. lipolytica* to produce medium-chain fatty acids with high amounts, in agreement with the work previously [189]. However, the earlier worked did not investigated variations in medium-chain fatty acids production among the clones with the same acyl-ACP thioesterase due to the effect of random integration. Random integration uses a non-homologous end-joining mechanism to introduce heterologous gene fragments. Compared to targeted integration using CRISPR/Cas9 system, it has a much higher integration efficiency, but can lead to variation in gene location and copy number. All four strains showed different integrated sites in the chromosome. One of the strains (Clone #19) contained multiple gene copies of acyl-ACP thioesterase, while the others contained one gene copy. Homology-independent random genome integration created variation in protein expression levels [199]. As a result, different clones showed different acyl-ACP thioesterase expression and ability to produce medium-chain fatty acids. The phenotypical variation caused by homology-

independent integration allows us to screen for the most optimal lipid producer and is more flexible for multiple metabolic pathway construction compared to targeted integration.

It is unclear how a plant or bacterial acyl-ACP thioesterase associates with the fungal type I FAS because most of the pertinent catalytic sites reside in the interior of the capsule-like structure [205,206]. A heterologous acyl-ACP thioesterase may disrupt the natural structure of the FAS to allow accessibility of the thioesterase catalytic sites to the growing acyl-ACP chains during elongation [189]. We found that even with a fairly high expression of acyl-ACP thioesterase in Clone #1, there is little production of medium-chain fatty acids, suggesting that association of plant thioesterase and type I FAS may not be unfavorable and need a even higher concentration of acyl-ACP thioesterase. In general, it is considered that multiple copies of a gene result in a higher expression compared to a single copy. However, acyl-ACP thioesterase expression in Clone #20 with one copy exceeded the expression in Clone #19 with three copies, also reflected by the higher titer in C8 fatty acids production in Clone #20.

Using whole genome sequencing and genome walking kit, we were able to determine the gene location and copy number of random integrated DNA fragments. However, genome walking does not guarantee to find locations for all possible gene copies, and quantitative PCR can be used as a supplementary method. The number of gene copies can be estimated using a pair of primers for the integrated region, and another pair for the reference genome. The ratio for the respective amplicon can be used to determine copy numbers. However, high GC ratio and multiple copies of the gene can pose problems for genome walking. It is possible to find mixed PCR fragments difficult to separate and identify.

Chapter 6: Outlook

Oleaginous yeasts are an emerging host for lignocellulosic biomass for the production of biofuels and bioproducts. Research on *Rhodosporidium toruloides* and *Lipomyces starkeyi* has highlighted their potential for growth on complex substrates present in lignocellulosic hydrolysate and their tolerance to the inhibitors. However, in comparison to model organisms like *Saccharomyces cerevisiae*, our understanding of these yeasts is pretty limited. They were sequenced in the last decade[10,11], and recent advancements in the Next Generation Sequencing technologies have provided us with more accurate genome sequences[41].

This work has attempted to improve the understanding of both these yeasts. We have analysed the growth of *R. toruloides* and *L. starkeyi* on various plant based sugars and oils and mapped out the pathways involved in the utilization of these sugars. We have also improved the gene annotation of both yeasts and created public datasets with the improved annotations[127]. These datasets will improve our understanding of the genomics of these yeasts. We have also identified a monosaccharide transporter from *L.starkeyi* capable of uninhibited co-consumption of glucose and xylose[120]. In addition, we have generated pipelines for both DNA and RNA-sequencing data analysis for *R. toruloides*, *L. starkeyi,* and *Y. lipolytica*, which can be directly used for analysis in future projects.

While we have attempted to improve our understanding of these yeasts, we have merely scratched the surface. Future studies focused on mixed substrate utilization in the yeasts, and the effect of the inhibitors present in lignocellulosic biomass will present a better understanding of the fermentation conditions. In the last 5 years of research on *R. toruloides*, dozens of transcriptomics datasets have been generated. These datasets can be processed to generate genome-wide insights, understand metabolic regulation, and generate gene regulatory networks.

Lipid biosynthesis in oleaginous yeasts is an interesting metabolic phenomenon that shares similarities with the human lipid pathways. A detailed understanding of the mechanisms and regulations of this phenomenon will help with the rational design of high lipid-producing strains.

References

1. Preiss, J. Bacterial Glycogen Synthesis and its Regulation. *Annu. Rev. Microbiol.* **38**, 419–458 (1984).
2. Athenaki, M. *et al.* Lipids from yeasts and fungi: physiology, production and analytical considerations. *J. Appl. Microbiol.* **124**, 336–367 (2018).
3. Bhutada, G. *et al.* Sugar versus fat: Elimination of glycogen storage improves lipid accumulation in Yarrowia lipolytica. *FEMS Yeast Res.* **17**, 1–10 (2017).
4. Graef, M. Lipid droplet-mediated lipid and protein homeostasis in budding yeast. *FEBS Lett.* **592**, 1291–1303 (2018).
5. Yang, Y. & Hu, B. Bio-based chemicals from biorefining: lipid and wax conversion and utilization. in *Advances in Biorefineries* 693–720 (Woodhead Publishing, 2014). doi:10.1533/9780857097385.2.693.
6. Okolie, J. A., Nanda, S., Dalai, A. K. & Kozinski, J. A. Chemistry and Specialty Industrial Applications of Lignocellulosic Biomass. *Waste and Biomass Valorization* **12**, 2145–2169 (2021).
7. Werpy, T. & Petersen, G. Top Value Added Chemicals from Biomass Volume I. *Us Nrel* Medium: ED; Size: 76 pp. pages (2004) doi:10.2172/15008859.
8. Beopoulos, A. *et al.* Control of lipid accumulation in the yeast Yarrowia lipolytica. *Appl. Environ. Microbiol.* **74**, 7779–89 (2008).
9. Meng, X. *et al.* Biodiesel production from oleaginous microorganisms. *Renew. Energy* **34**, 1–5 (2009).
10. Zhang, S. *et al.* Engineering Rhodosporidium toruloides for increased lipid production. *Biotechnol Bioeng* **113**, 1056–1066 (2016).
11. Riley, R. *et al.* Comparative genomics of biotechnologically important yeasts. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 9882–7 (2016).
12. Toivari, M. H., Salusjärvi, L., Ruohonen, L. & Penttilä, M. Endogenous xylose pathway in Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* **70**, 3681–6 (2004).
13. Zhu, Z. *et al.* A multi-omic map of the lipid-producing yeast Rhodosporidium toruloides. *Nat. Commun.* **3**, 1112 (2012).
14. Wang, Y. *et al.* Systems analysis of phosphate-limitation-induced lipid accumulation by the oleaginous yeast Rhodosporidium toruloides. *Biotechnol Biofuels* **11**, 148 (2018).
15. Coradetti, S. T. *et al.* Functional genomics of lipid metabolism in the oleaginous yeast Rhodosporidium toruloides. *Elife* **7**, 1–55 (2018).
16. Fei, Q. *et al.* Enhanced lipid production by Rhodosporidium toruloides using different fed-batch feeding strategies with lignocellulosic hydrolysate as the sole carbon source. *Biotechnol Biofuel* **9**, 130 (2016).
17. Hu, C., Zhao, X., Zhao, J., Wu, S. & Zhao, Z. K. Effects of biomass hydrolysis by-products on oleaginous yeast Rhodosporidium toruloides. *Bioresour Technol* **100**, 4843–4847 (2009).
18. Huang, X.-F. *et al.* Culture strategies for lipid production using acetic acid as sole carbon

source by Rhodosporidium toruloides. *Bioresour Technol* **206**, 141–149 (2016).

19. Kitahara, Y. *et al.* Isolation of oleaginous yeast (Rhodosporidium toruloides) mutants tolerant of sugarcane bagasse hydrolysate. *Biosci Biotechnol Biochem* **78**, 336–342 (2014).

20. Singh, G. *et al.* Concomitant production of lipids and carotenoids in Rhodosporidium toruloides under osmotic stress using response surface methodology. *Front Microbiol* **7**, (2016).

21. Wiebe, M. G., Koivuranta, K., Penttilä, M. & Ruohonen, L. Lipid production in batch and fed-batch cultures of Rhodosporidium toruloides from 5 and 6 carbon carbohydrates. *BMC Biotechnol* **12**, 26 (2012).

22. Yaegashi, J. *et al.* Rhodosporidium toruloides: a new platform organism for conversion of lignocellulose into terpene biofuels and bioproducts. *Biotechnol Biofuels* **10**, 241 (2017).

23. Zhang, S., Ito, M., Skerker, J. M., Arkin, A. P. & Rao, C. V. Metabolic engineering of the oleaginous yeast Rhodosporidium toruloides IFO0880 for lipid overproduction during high-density fermentation. *Appl. Microbiol. Biotechnol.* **100**, 9393–9405 (2016).

24. Wu, S., Zhao, X., Shen, H., Wang, Q. & Zhao, Z. K. Microbial lipid production by Rhodosporidium toruloides under sulfate-limited conditions. *Bioresour Technol* **102**, 1803–1807 (2011).

25. Jagtap, S. S. & Rao, C. V. Microbial conversion of xylose into useful bioproducts. *Appl. Microbiol. Biotechnol.* **102**, 9015–9036 (2018).

26. Dinh, H. V. *et al.* A comprehensive genome-scale model for Rhodosporidium toruloides IFO0880 accounting for functional genomics and phenotypic data. **9**, (2019).

27. Fillet, S. *et al.* Fatty alcohols production by oleaginous yeast. *J Ind Microbiol Biotechnol* **42**, 1463–1472 (2015).

28. Fillet, S. *et al.* Engineering Rhodosporidium toruloides for the production of very long-chain monounsaturated fatty acid-rich oils. *Appl Microbiol Biotechnol* **101**, 7271–7280 (2017).

29. Liu, D. *et al.* Exploiting nonionic surfactants to enhance fatty alcohol production in Rhodosporidium toruloides. *Biotechnol Bioeng* **117**, 1418–1425 (2020).

30. Lee, J. J. L., Chen, L., Shi, J., Trzcinski, A. & Chen, W.-N. Metabolomic profiling of Rhodosporidium toruloides grown on glycerol for carotenoid production during different growth phases. *J Agric Food Chem* **62**, 10203–10209 (2014).

31. Lee, J. J. L., Chen, L., Cao, B. & Chen, W. N. Engineering Rhodosporidium toruloides with a membrane transporter facilitates production and separation of carotenoids and lipids in a bi-phasic culture. *Appl Microbiol Biotechnol* **100**, 869–877 (2016).

32. Wehrs, M. *et al.* Sustainable bioproduction of the blue pigment indigoidine: Expanding the range of heterologous products in R. toruloides to include non-ribosomal peptides. *Green Chem* **21**, 3394–3406 (2019).

33. Wen, Z., Zhang, S., Odoh, C. K., Jin, M. & Zhao, Z. K. Rhodosporidium toruloides - A potential red yeast chassis for lipids and beyond. *FEMS Yeast Res.* **20**, 38 (2020).

34. Zhang, S. *et al.* Engineering *Rhodosporidium toruloides* for increased lipid production.

Biotechnol. Bioeng. **113**, 1056–1066 (2016).

35. Cui, J. D., Qiu, J. Q., Fan, X. W., Jia, S. R. & Tan, Z. L. Biotechnological production and applications of microbial phenylalanine ammonia lyase: a recent review. *Crit Rev Biotechnol* **34**, 258–268 (2014).

36. Fernández-Lafuente, R., Rodriguez, V. & Guisán, J. M. The coimmobilization of d-amino acid oxidase and catalase enables the quantitative transformation of d-amino acids (d-phenylalanine) into α-keto acids (phenylpyruvic acid). *Enzym. Microb Technol* **23**, 28–33 (1998).

37. Gilbert, H. J., Clarke, I. N., Gibson, R. K., Stephenson, J. R. & Tully, M. Molecular cloning of the phenylalanine ammonia lyase gene from Rhodosporidium toruloides in Escherichia coli K-12. *J Bacteriol* **161**, 314–320 (1985).

38. Hoskins, J. *et al.* Enzymatic control of phenylalanine intake in phenylketonuria. *Lancet* **315**, 392–394 (1980).

39. Pilone, M. S. & Pollegioni, L. D-amino acid oxidase as an industrial biocatalyst. *Biocatal Biotransfor* **20**, 145–159 (2002).

40. Zhu, Z. *et al.* Dynamics of the lipid droplet proteome of the oleaginous yeast Rhodosporidium toruloides. *Eukaryot. Cell* **14**, 252–264 (2015).

41. Coradetti, S. T. *et al.* Functional genomics of lipid metabolism in the oleaginous yeast Rhodosporidium toruloides. *Elife* **7**, e32110 (2018).

42. Nora, L. C. *et al.* A toolset of constitutive promoters for metabolic engineering of Rhodosporidium toruloides. *Microb Cell Fact* **18**, 117 (2019).

43. Kim, J. *et al.* Multi-Omics Driven Metabolic Network Reconstruction and Analysis of Lignocellulosic Carbon Utilization in Rhodosporidium toruloides. *Front. Bioeng. Biotechnol.* **8**, (2021).

44. Zhang, S., Jagtap, S. S., Deewan, A. & Rao, C. V. pH selectively regulates citric acid and lipid production in Yarrowia lipolytica W29 during nitrogen-limited growth on glucose. *J. Biotechnol.* **290**, (2019).

45. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

46. Andrews, S. FastQC: a quality control tool for high throughput sequence data. (2010).

47. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

48. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

49. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).

50. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

51. Blighe K, L. A. PCAtools: everything principal components analysis. *R Packag. version 2.4.0* (2021).

52. Su, S. *et al.* Glimma: interactive graphics for gene expression analysis. *Bioinformatics* **33**, 2050–2052 (2017).

53. Warnes, G. R. Gplots: Various R Programming Tools for Plotting Data. (2011).

54. Grigoriev, I. V *et al.* MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res* **42**, D699-704 (2014).

55. Kim, S. *et al.* Evaluation and optimization of metabolome sample preparation methods for Saccharomyces cerevisiae. *Anal Chem* **85**, 2169–2176 (2013).

56. Yun, E. J. *et al.* Metabolomic elucidation of the effects of media and carbon sources on fatty acid production by Yarrowia lipolytica. *J. Biotechnol.* **272**–**273**, 7–13 (2018).

57. Stein, S. E. An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J Am Soc Mass Spectrom* **10**, 770–781 (1999).

58. Kopka, J. *et al.* GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics* **21**, 1635–1638 (2005).

59. Styczynski, M. P. *et al.* Systematic identification of conserved metabolites in GC/MS data for metabolomics and biomarker discovery. *Anal Chem* **79**, 966–973 (2007).

60. Howe, E. *et al.* MeV: MultiExperiment Viewer. in *Biomedical Informatics for Cancer Research* (eds. Ochs, M. F., Casagrande, J. T. & Davuluri, R. V) 267–277 (Springer US, 2010). doi:10.1007/978-1-4419-5714-6_15.

61. Chong, J., Wishart, D. S. & Xia, J. Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis. *Curr Protoc Bioinforma.* **68**, e86 (2019).

62. Lutfiyya, L. L. & Johnston, M. Two zinc-finger-containing repressors are responsible for glucose repression of SUC2 expression. *Mol Cell Biol* **16**, 4790–4797 (1996).

63. Mormeneo, S. & Sentandreu, R. Regulation of invertase synthesis by glucose in Saccharomyces cerevisiae. *J Bacteriol* **152**, 14–18 (1982).

64. Jagtap, S. S. & Rao, C. V. Production of D-arabitol from D-xylose by the oleaginous yeast Rhodosporidium toruloides IFO0880. *Appl. Microbiol. Biotechnol.* **102**, 143–151 (2018).

65. Ingram, J. M. & Wood, W. A. Enzymatic basis for D-arbitol production by Zygosaccharomyces rouxii. *J Bacteriol* **89**, 1186–1194 (1965).

66. Wong, B., Murray, J. S., Castellanos, M. & Croen, K. D. D-arabitol metabolism in Candida albicans: studies of the biosynthetic pathway and the gene that encodes NAD-dependent D-arabitol dehydrogenase. *J Bacteriol* **175**, 6314–6320 (1993).

67. Toivari, M. H., Nygård, Y., Penttilä, M., Ruohonen, L. & Wiebe, M. G. Microbial D-xylonate production. *Appl Microbiol Biotechnol* **96**, 1–8 (2012).

68. Casal, M., Paiva, S., Queiros, O. & Soares-Silva, I. Transport of carboxylic acids in yeasts. *FEMS Microbiol Rev* **32**, 974–994 (2008).

69. Fickers, P., Marty, A. & Nicaud, J. M. The lipases from Yarrowia lipolytica: genetics, production, regulation, biochemical characterization and biotechnological applications. *Biotechnol. Adv.* **29**, 632–44 (2011).

70. Færgeman, N. J., Black, P. N., Zhao, X. D., Knudsen, J. & DiRusso, C. C. The Acyl-CoA

synthetases encoded within FAA1 and FAA4 in Saccharomyces cerevisiae function as components of the fatty acid transport system linking import, activation, and intracellular utilization. *J Biol Chem* **276**, 37051–37059 (2001).

71. Johnson, D. R., Knoll, L. J., Levin, D. E. & Gordon, J. I. Saccharomyces cerevisiae contains four fatty acid activation (FAA) genes: an assessment of their role in regulating protein N-myristoylation and cellular lipid metabolism. *J Cell Biol* **127**, 751–762 (1994).

72. Hiltunen, J. K. *et al.* The biochemistry of peroxisomal beta-oxidation in the yeast Saccharomyces cerevisiae. *FEMS Microbiol Rev* **27**, 35–64 (2003).

73. Wang, H. J. *et al.* Evaluation of acyl coenzyme A oxidase (Aox) isozyme function in the n-alkane-assimilating yeast Yarrowia lipolytica. *J Bacteriol* **181**, 5140–5148 (1999).

74. Wang, H. *et al.* Cloning, sequencing, and characterization of five genes coding for acyl-CoA oxidase isozymes in the yeast Yarrowia lipolytica. *Cell Biochem Biophys* **31**, 165–174 (1999).

75. Hiltunen, J. K. *et al.* Peroxisomal multifunctional beta-oxidation protein of Saccharomyces cerevisiae. Molecular analysis of the fox2 gene and gene product. *J Biol Chem* **267**, 6646–6653 (1992).

76. Igual, J. C., Matallana, E., Gonzalez-Bosch, C., Franco, L. & Perez-Ortin, J. E. A new glucose-repressible gene identified from the analysis of chromatin structure in deletion mutants of yeast SUC2 locus. *Yeast* **7**, 379–389 (1991).

77. Elgersma, Y., van Roermund, C. W., Wanders, R. J. & Tabak, H. F. Peroxisomal and mitochondrial carnitine acetyltransferases of Saccharomyces cerevisiae are encoded by a single gene. *EMBO J* **14**, 3472–3479 (1995).

78. Strijbis, K. & Distel, B. Intracellular acetyl unit transport in fungal carbon metabolism. *Eukaryot Cell* **9**, 1809–1815 (2010).

79. Swiegers, J. H., Dippenaar, N., Pretorius, I. S. & Bauer, F. F. Carnitine-dependent metabolic activities in Saccharomyces cerevisiae: three carnitine acetyltransferases are essential in a carnitine-dependent strain. *Yeast* **18**, 585–595 (2001).

80. Palmieri, L. *et al.* Identification of the mitochondrial carnitine carrier in Saccharomyces cerevisiae. *FEBS Lett* **462**, 472–476 (1999).

81. Pinheiro, M. J., Bonturi, N., Belouah, I., Miranda, E. A. & Lahtvee, P.-J. Xylose Metabolism and the Effect of Oxidative Stress on Lipid and Carotenoid Production in Rhodotorula toruloides: Insights for Future Biorefinery. *Front. Bioeng. Biotechnol.* **8**, 1008 (2020).

82. Quarterman, J., Slininger, P. J., Kurtzman, C. P., Thompson, S. R. & Dien, B. S. A survey of yeast from the Yarrowia clade for lipid production in dilute acid pretreated lignocellulosic biomass hydrolysate. *Appl Microbiol Biotechnol* **101**, 3319–3334 (2017).

83. Buu, L. M., Chen, Y. C. & Lee, F. J. Functional characterization and localization of acetyl-CoA hydrolase, Ach1p, in Saccharomyces cerevisiae. *J Biol Chem* **278**, 17203–17209 (2003).

84. Fleck, C. B. & Brock, M. Re-characterisation of Saccharomyces cerevisiae Ach1p: fungal CoA-transferases are involved in acetic acid detoxification. *Fungal Genet Biol* **46**, 473–485 (2009).

85. Choi, J.-Y. & Martin, C. E. The Saccharomyces cerevisiae FAT1 gene encodes an acyl-CoA synthetase that is required for maintenance of very long chain fatty acid levels. *J Biol Chem* **274**, 4671–4683 (1999).

86. Watkins, P. A. *et al.* Disruption of the Saccharomyces cerevisiae FAT1 gene decreases very long-chain fatty acyl-coa synthetase activity and elevates intracellular very long-chain fatty acid concentrations. *J Biol Chem* **273**, 18210–18219 (1998).

87. Watkins, P. A. Very-long-chain acyl-CoA synthetases. *J Biol Chem* **283**, 1773–1777 (2008).

88. Watkins, P. A., Maiguel, D., Jia, Z. & Pevsner, J. Evidence for 26 distinct acyl-coenzyme A synthetase genes in the human genome. *J Lipid Res* **48**, 2736–2750 (2007).

89. Coe, N. R., Smith, A. J., Frohnert, B. I., Watkins, P. A. & Bernlohr, D. A. The fatty acid transport protein (FATP1) is a very long chain acyl-CoA synthetase. *J Biol Chem* **274**, 36300–36304 (1999).

90. Chegary, M. *et al.* Mitochondrial long chain fatty acid β-oxidation in man and mouse. *Biochim. Biophys. Acta* **1791**, 806–815 (2009).

91. Swigonová, Z., Mohsen, A. W. & Vockley, J. Acyl-CoA dehydrogenases: Dynamic history of protein family evolution. *J Mol Evol* **69**, 176–193 (2009).

92. Castañeda, M. T., Nuñez, S., Garelli, F., Voget, C. & De Battista, H. Comprehensive analysis of a metabolic model for lipid production in Rhodosporidium toruloides. *J Biotechnol* **280**, 11–18 (2018).

93. Tiukova, I. A., Prigent, S., Nielsen, J., Sandgren, M. & Kerkhoven, E. J. Genome-scale model of Rhodotorula toruloides metabolism. *Biotechnol Bioeng* **116**, 3396–3408 (2019).

94. Ageitos, J. M., Vallejo, J. A., Veiga-Crespo, P. & Villa, T. G. Oily yeasts as oleaginous cell factories. *Appl Microbiol Biotechnol* **90**, 1219–1227 (2011).

95. Shi, S. & Zhao, H. Metabolic Engineering of Oleaginous Yeasts for Production of Fuels and Chemicals. *Front Microbiol* **8**, 2185 (2017).

96. Spagnuolo, M., Yaguchi, A. & Blenner, M. Oleaginous yeast for biofuel and oleochemical production. *Curr Opin Biotechnol* **57**, 73–81 (2019).

97. Starkey, R. L. Lipid production by a soil yeast. *J Bacteriol* **51**, 33–50 (1946).

98. Lodder, J. & Kreger-van Rij, N. J. W. Genus Lipomyces. in *The yeasts, a taxonomic study* (eds. Lodder, J. & Kreger-van Rij, N. J. W.) 669–700 (North-Holland Publishing Company (Elsevier), 1952).

99. McNeil, B. A. & Stuart, D. T. Lipomyces starkeyi: an emerging cell factory for production of lipids, oleochemicals and biotechnology applications. *World J Microbiol Biotechnol* **34**, 147 (2018).

100. Bao, M. *et al.* Identification, soluble expression, and characterization of a novel endo-inulinase from Lipomyces starkeyi NRRL Y-11557. *Int J Biol Macromol* **137**, 537–544 (2019).

101. Kang, H. K. *et al.* Cloning and expression of Lipomyces starkeyi alpha-amylase in Escherichia coli and determination of some of its properties. *FEMS Microbiol Lett* **233**, 53–64 (2004).

102. Kang, H. K., Park, J. Y., Ahn, J. S., Kim, S. H. & Kim, D. Cloning of a gene encoding dextranase from Lipomyces starkeyi and its expression in Pichia pastoris. *J Microbiol Biotechnol* **19**, 172–177 (2009).

103. Nishimura, K. *et al.* Identification of enzyme responsible for erythritol utilization and reaction product in yeast Lipomyces starkeyi. *J Biosci Bioeng* **101**, 303–308 (2006).

104. Rother, C. *et al.* Biochemical Characterization and Mechanistic Analysis of the Levoglucosan Kinase from Lipomyces starkeyi. *Chembiochem* **19**, 596–603 (2018).

105. Ryu, S. J. *et al.* Purification and partial characterization of a novel glucanhydrolase from Lipomyces starkeyi KSM 22 and its use for inhibition of insoluble glucan formation. *Biosci Biotechnol Biochem* **64**, 223–228 (2000).

106. Calvey, C. H., Su, Y.-K. K., Willis, L. B., McGee, M. & Jeffries, T. W. Nitrogen limitation, oxygen limitation, and lipid accumulation in Lipomyces starkeyi. *Bioresour Technol* **200**, 780–788 (2016).

107. Juanssilfero, A. B. *et al.* Effect of inoculum size on single-cell oil production from glucose and xylose using oleaginous yeast Lipomyces starkeyi. *J Biosci Bioeng* **125**, 695–702 (2018).

108. Kamineni, A. & Shaw, J. Engineering triacylglycerol production from sugars in oleaginous yeasts. *Curr Opin Biotechnol* **62**, 239–247 (2020).

109. Takaku, H., Matsuzawa, T., Yaoi, K. & Yamazaki, H. Lipid metabolism of the oleaginous yeast Lipomyces starkeyi. *Appl Microbiol Biotechnol* **104**, 6141–6148 (2020).

110. McNeil, B. A. & Stuart, D. T. Optimization of C16 and C18 fatty alcohol production by an engineered strain of Lipomyces starkeyi. *J Ind Microbiol Biotechnol* **45**, 1–14 (2018).

111. Wang, W. *et al.* Fatty alcohol production in Lipomyces starkeyi and Yarrowia lipolytica. *Biotechnol Biofuels* **9**, 227 (2016).

112. Lin, J. *et al.* Lipid production by Lipomyces starkeyi cells in glucose solution without auxiliary nutrients. *J Biotechnol* **152**, 184–188 (2011).

113. Zhang, L., Lee, J. T. E., Ok, Y. S., Dai, Y. & Tong, Y. W. Enhancing microbial lipids yield for biodiesel production by oleaginous yeast Lipomyces starkeyi fermentation: A review. *Bioresour Technol* **344**, 126294 (2022).

114. Zhang, L., Lim, E. Y., Loh, K. C., Dai, Y. & Tong, Y. W. Two-Stage Fermentation of Lipomyces starkeyi for Production of Microbial Lipids and Biodiesel. *Microorganisms* **9**, (2021).

115. Monteiro de Oliveira, P., Aborneva, D., Bonturi, N. & Lahtvee, P. J. Screening and Growth Characterization of Non-conventional Yeasts in a Hemicellulosic Hydrolysate. *Front Bioeng Biotechnol* **9**, 659472 (2021).

116. Pomraning, K. R. *et al.* Transcriptomic analysis of the oleaginous yeast Lipomyces starkeyi during lipid accumulation on enzymatically treated corn stover hydrolysate. *Biotechnol Biofuels* **12**, 162 (2019).

117. Riley, R. *et al.* Comparative genomics of biotechnologically important yeasts. *Proc Natl Acad Sci U S A* **113**, 9882–9887 (2016).

118. Zhou, W. *et al.* A metabolic model of Lipomyces starkeyi for predicting lipogenesis

potential from diverse low-cost substrates. *Biotechnol Biofuels* **14**, 148 (2021).

119. de Ruijter, J. C., Igarashi, K., Penttilä, M. & Penttila, M. The Lipomyces starkeyi gene Ls120451 encodes a cellobiose transporter that enables cellobiose fermentation in Saccharomyces cerevisiae. *FEMS Yeast Res* **20**, 1–13 (2020).

120. Kuanyshev, N. *et al.* Identification and analysis of sugar transporters capable of co-transporting glucose and xylose simultaneously. *Biotechnol J* **16**, e2100238 (2021).

121. Aburatani, S. *et al.* Inference of Regulatory System for TAG Biosynthesis in Lipomyces starkeyi. *Bioeng.* **7**, 1–15 (2020).

122. Calvey, C. H., Willis, L. B. & Jeffries, T. W. An optimized transformation protocol for Lipomyces starkeyi. *Curr Genet* **60**, 223–230 (2014).

123. Dai, Z., Deng, S., Culley, D. E., Bruno, K. S. & Magnuson, J. K. Agrobacterium tumefaciens-mediated transformation of oleaginous yeast Lipomyces species. *Appl Microbiol Biotechnol* **101**, 6099–6110 (2017).

124. Lin, X. *et al.* Development of an Agrobacterium-Mediated Transformation Method and Evaluation of Two Exogenous Constitutive Promoters in Oleaginous Yeast Lipomyces starkeyi. *Appl Biochem Biotechnol* **183**, 867–875 (2017).

125. Takaku, H. *et al.* A novel electroporation procedure for highly efficient transformation of Lipomyces starkeyi. *J Microbiol Methods* **169**, 105816 (2020).

126. Oguro, Y. *et al.* Efficient gene targeting in non-homologous end-joining-deficient Lipomyces starkeyi strains. *Curr Genet* **63**, 751–763 (2017).

127. Jagtap, S. S. *et al.* Integrating transcriptomic and metabolomic analysis of the oleaginous yeast Rhodosporidium toruloides IFO0880 during growth under different carbon sources. *Appl Microbiol Biotechnol* **105**, 7411–7425 (2021).

128. Liu, J. J. *et al.* Investigating the role of the transcriptional regulator Ure2 on the metabolism of Saccharomyces cerevisiae: a multi-omics approach. *Appl Microbiol Biotechnol* **105**, 5103–5112 (2021).

129. Andrews, S. *et al.* FastQC. (2012).

130. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545–15550 (2005).

131. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498–2504 (2003).

132. Almagro Armenteros, J. J., Sonderby, C. K., Sonderby, S. K., Nielsen, H. & Winther, O. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* **33**, 3387–3395 (2017).

133. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27–30 (2000).

134. Lee, J. W., Yook, S., Koh, H., Rao, C. V & Jin, Y. S. Engineering xylose metabolism in yeasts to produce biofuels and chemicals. *Curr Opin Biotechnol* **67**, 15–25 (2021).

135. Gong, Z. *et al.* Co-fermentation of cellobiose and xylose by Lipomyces starkeyi for lipid production. *Bioresour Technol* **117**, 20–24 (2012).

136. Brandenburg, J. *et al.* Oleaginous yeasts respond differently to carbon sources present in lignocellulose hydrolysate. *Biotechnol Biofuels* **14**, 124 (2021).

137. Ha, S. J. *et al.* Engineered Saccharomyces cerevisiae capable of simultaneous cellobiose and xylose fermentation. *Proc Natl Acad Sci U S A* **108**, 504–509 (2011).

138. Kwak, S., Jo, J. H., Yun, E. J., Jin, Y. S. & Seo, J. H. Production of biofuels and chemicals from xylose using native and engineered yeast strains. *Biotechnol Adv* **37**, 271–283 (2019).

139. Carroll, A. & Somerville, C. Cellulosic biofuels. *Annu Rev Plant Biol* **60**, 165–182 (2009).

140. Kim, S. R., Ha, S. J., Wei, N., Oh, E. J. & Jin, Y. S. Simultaneous co-fermentation of mixed sugars: a promising strategy for producing cellulosic ethanol. *Trends Biotechnol* **30**, 274–282 (2012).

141. Brat, D., Boles, E. & Wiedemann, B. Functional expression of a bacterial xylose isomerase in Saccharomyces cerevisiae. *Appl Env. Microbiol* **75**, 2304–2311 (2009).

142. Jeffries, T. W. & Jin, Y. S. Metabolic engineering for improved fermentation of pentoses by yeasts. *Appl Microbiol Biotechnol* **63**, 495–509 (2004).

143. Jin, Y. S., Ni, H., Laplaza, J. M. & Jeffries, T. W. Optimal growth and ethanol production from xylose by recombinant Saccharomyces cerevisiae require moderate D-xylulokinase activity. *Appl Env. Microbiol* **69**, 495–503 (2003).

144. Kim, S. R. *et al.* Rational and evolutionary engineering approaches uncover a small set of genetic changes efficient for rapid xylose fermentation in Saccharomyces cerevisiae. *PLoS One* **8**, e57048 (2013).

145. Kotter, P., Amore, R., Hollenberg, C. P. & Ciriacy, M. Isolation and characterization of the Pichia stipitis xylitol dehydrogenase gene, XYL2, and construction of a xylose-utilizing Saccharomyces cerevisiae transformant. *Curr Genet* **18**, 493–500 (1990).

146. Kuyper, M. *et al.* High-level functional expression of a fungal xylose isomerase: the key to efficient ethanolic fermentation of xylose by Saccharomyces cerevisiae? *FEMS Yeast Res* **4**, 69–78 (2003).

147. Kwak, S. & Jin, Y. S. Production of fuels and chemicals from xylose by engineered Saccharomyces cerevisiae: a review and perspective. *Microb Cell Fact* **16**, 82 (2017).

148. Zhou, H., Cheng, J. S., Wang, B. L., Fink, G. R. & Stephanopoulos, G. Xylose isomerase overexpression along with engineering of the pentose phosphate pathway and evolutionary engineering enable rapid xylose utilization and ethanol production by Saccharomyces cerevisiae. *Metab Eng* **14**, 611–622 (2012).

149. Gardonyi, M., Jeppsson, M., Liden, G., Gorwa-Grauslund, M. F. & Hahn-Hagerdal, B. Control of xylose consumption by xylose transport in recombinant Saccharomyces cerevisiae. *Biotechnol Bioeng* **82**, 818–824 (2003).

150. Hamacher, T., Becker, J., Gardonyi, M., Hahn-Hagerdal, B. & Boles, E. Characterization of the xylose-transporting properties of yeast hexose transporters and their influence on xylose utilization. *Microbiology* **148**, 2783–2788 (2002).

151. Parachin, N. S., Bergdahl, B., van Niel, E. W. J. & Gorwa-Grauslund, M. F. Kinetic modelling reveals current limitations in the production of ethanol from xylose by

recombinant Saccharomyces cerevisiae. *Metab. Eng.* **13**, 508–517 (2011).

152. Sedlak, M. & Ho, N. W. Characterization of the effectiveness of hexose transporters for transporting xylose during glucose and xylose co-fermentation by a recombinant Saccharomyces yeast. *Yeast* **21**, 671–684 (2004).

153. Ozcan, S. & Johnston, M. Function and regulation of yeast hexose transporters. *Microbiol. Mol. Biol. Rev.* **63**, 554–69 (1999).

154. Subtil, T. & Boles, E. Competition between pentoses and glucose during uptake and catabolism in recombinant Saccharomyces cerevisiae. *Biotechnol Biofuels* **5**, 14 (2012).

155. Cheng, M. H., Dien, B. S., Lee, D. K. & Singh, V. Sugar production from bioenergy sorghum by using pilot scale continuous hydrothermal pretreatment combined with disk refining. *Bioresour Technol* **289**, 121663 (2019).

156. Shirkavand, E., Baroutian, S., Gapes, D. J. & Young, B. R. Combination of fungal and physicochemical processes for lignocellulosic biomass pretreatment – A review. *Renew. Sustain. Energy Rev.* **54**, 217–234 (2016).

157. Leandro, M. J., Fonseca, C. & Goncalves, P. Hexose and pentose transport in ascomycetous yeasts: an overview. *FEMS Yeast Res.* **9**, 511–525 (2009).

158. Young, E., Poucher, A., Comer, A., Bailey, A. & Alper, H. Functional survey for heterologous sugar transport proteins, using Saccharomyces cerevisiae as a host. *Appl. Environ. Microbiol.* **77**, 3311–9 (2011).

159. Farwick, A., Bruder, S., Schadeweg, V., Oreb, M. & Boles, E. Engineering of yeast hexose transporters to transport D-xylose without inhibition by D-glucose. *Proc Natl Acad Sci U S A* **111**, 5159–5164 (2014).

160. Li, H., Schmitz, O. & Alper, H. S. Enabling glucose/xylose co-transport in yeast through the directed evolution of a sugar transporter. *Appl Microbiol Biotechnol* **100**, 10215–10223 (2016).

161. Reider Apel, A., Ouellet, M., Szmidt-Middleton, H., Keasling, J. D. & Mukhopadhyay, A. Evolved hexose transporter enhances xylose uptake and glucose/xylose co-utilization in Saccharomyces cerevisiae. *Sci Rep* **6**, 19512 (2016).

162. Shin, H. Y. *et al.* An engineered cryptic Hxt11 sugar transporter facilitates glucose-xylose co-consumption in Saccharomyces cerevisiae. *Biotechnol Biofuels* **8**, 176 (2015).

163. Young, E. M., Tong, A., Bui, H., Spofford, C. & Alper, H. S. Rewiring yeast sugar transporter preference through modifying a conserved protein motif. *Proc Natl Acad Sci U S A* **111**, 131–136 (2014).

164. Jeena, G. S., Kumar, S. & Shukla, R. K. Structure, evolution and diverse physiological roles of SWEET sugar transporters in plants. *Plant Mol Biol* **100**, 351–365 (2019).

165. Chen, L. Q. *et al.* Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature* **468**, 527–532 (2010).

166. Han, L. *et al.* Molecular mechanism of substrate recognition and transport by the AtSWEET13 sugar transporter. *Proc Natl Acad Sci U S A* **114**, 10089–10094 (2017).

167. Tao, Y. *et al.* Structure of a eukaryotic SWEET transporter in a homotrimeric complex. *Nature* **527**, 259–263 (2015).

168. Xuan, Y. H. *et al.* Functional role of oligomerization for bacterial and plant SWEET sugar transporter family. *Proc Natl Acad Sci U S A* **110**, E3685-94 (2013).

169. Selvam, B., Yu, Y.-C., Chen, L.-Q. & Shukla, D. Molecular Basis of the Glucose Transport Mechanism in Plants. *ACS Cent. Sci.* acscentsci.9b00252 (2019) doi:10.1021/acscentsci.9b00252.

170. Podolsky, I. A., Seppala, S., Xu, H., Jin, Y. S. & O'Malley, M. A. A SWEET surprise: Anaerobic fungal sugar transporters and chimeras enhance sugar uptake in yeast. *Metab Eng* (2021) doi:10.1016/j.ymben.2021.04.009.

171. Xu, H. & Jin, Y.-S. Engineering Saccharomyces cerevisiae for cellulosic ethanol production. *Food Science & Human Nutrition* vol. M.S. (University of Illinois at Urbana-Champaign, 2015).

172. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc* **2**, 31–34 (2007).

173. Boles, E. & Oreb, M. A Growth-Based Screening System for Hexose Transporters in Yeast. *Methods Mol Biol* **1713**, 123–135 (2018).

174. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403–410 (1990).

175. Lewis, D. A. & Bisson, L. F. The HXT1 gene product of Saccharomyces cerevisiae is a new member of the family of hexose transporters. *Mol. Cell. Biol.* **11**, 3804–13 (1991).

176. Jeffries, T. W. *et al.* Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast Pichia stipitis. *Nat. Biotechnol.* **25**, 319–326 (2007).

177. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**, 1870–1874 (2016).

178. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

179. Cheng, K. J., Selvam, B., Chen, L. Q. & Shukla, D. Distinct Substrate Transport Mechanism Identified in Homologous Sugar Transporters. *J Phys Chem B* **123**, 8411–8418 (2019).

180. Eom, J. S. *et al.* SWEETs, transporters for intracellular and intercellular sugar translocation. *Curr Opin Plant Biol* **25**, 53–62 (2015).

181. Chen, L. Q., Cheung, L. S., Feng, L., Tanner, W. & Frommer, W. B. Transport of sugars. *Annu Rev Biochem* **84**, 865–894 (2015).

182. Fortman, J. L. *et al.* Biofuel alternatives to ethanol: pumping the microbial well. *Trends in Biotechnology* (2008) doi:10.1016/j.tibtech.2008.03.008.

183. Lazar, Z., Liu, N. & Stephanopoulos, G. Holistic Approaches in Lipid Production by Yarrowia lipolytica. *Trends in Biotechnology* (2018) doi:10.1016/j.tibtech.2018.06.007.

184. Nielsen, J., Larsson, C., van Maris, A. & Pronk, J. Metabolic engineering of yeast for production of fuels and chemicals. *Current Opinion in Biotechnology* (2013) doi:10.1016/j.copbio.2013.03.023.

185. Liu, H. *et al.* Fatty acid from the renewable sources: A promising feedstock for the

production of biofuels and biobased chemicals. *Biotechnology Advances* (2014) doi:10.1016/j.biotechadv.2013.12.003.

186. Runguphan, W. & Keasling, J. D. Metabolic engineering of Saccharomyces cerevisiae for production of fatty acid-derived biofuels and chemicals. *Metab. Eng.* (2014) doi:10.1016/j.ymben.2013.07.003.

187. Papanikolaou, S. & Aggelis, G. Lipids of oleaginous yeasts. Part II: Technology and potential applications. *European Journal of Lipid Science and Technology* (2011) doi:10.1002/ejlt.201100015.

188. Beopoulos, A. *et al.* Metabolic engineering for ricinoleic acid production in the oleaginous yeast Yarrowia lipolytica. *Appl. Microbiol. Biotechnol.* (2014) doi:10.1007/s00253-013-5295-x.

189. Rutter, C. D., Zhang, S. & Rao, C. V. Engineering Yarrowia lipolytica for production of medium-chain fatty acids. *Appl. Microbiol. Biotechnol.* **99**, 7359–7368 (2015).

190. Rigouin, C. *et al.* Increasing medium chain fatty acids production in Yarrowia lipolytica by metabolic engineering. *Microb. Cell Fact.* (2018) doi:10.1186/s12934-018-0989-5.

191. Ganesan, V. *et al.* Advances and opportunities in gene editing and gene regulation technology for Yarrowia lipolytica. *Microb. Cell Factories 2019 181* **18**, 1–9 (2019).

192. Zhu, Z. *et al.* Multidimensional engineering of Saccharomyces cerevisiae for efficient synthesis of medium-chain fatty acids. *Nat. Catal.* (2020) doi:10.1038/s41929-019-0409-1.

193. Sarria, S., Kruyer, N. S. & Peralta-Yahya, P. Microbial synthesis of medium-chain chemicals from renewables. *Nature Biotechnology* (2017) doi:10.1038/nbt.4022.

194. Yan, Q. & Pfleger, B. F. Revisiting metabolic engineering strategies for microbial synthesis of oleochemicals. *Metabolic Engineering* (2020) doi:10.1016/j.ymben.2019.04.009.

195. Zhu, Z. *et al.* Expanding the product portfolio of fungal type i fatty acid synthases. *Nat. Chem. Biol.* (2017) doi:10.1038/nchembio.2301.

196. Rigouin, C. *et al.* Production of Medium Chain Fatty Acids by Yarrowia lipolytica: Combining Molecular Design and TALEN to Engineer the Fatty Acid Synthase. *ACS Synth. Biol.* (2017) doi:10.1021/acssynbio.7b00034.

197. Xua, P., Qiao, K., Ahn, W. S. & Stephanopoulos, G. Engineering Yarrowia lipolytica as a platform for synthesis of drop-in transportation fuels and oleochemicals. *Proc. Natl. Acad. Sci. U. S. A.* (2016) doi:10.1073/pnas.1607295113.

198. Hernández Lozada, N. J. *et al.* Highly Active C 8 -Acyl-ACP Thioesterase Variant Isolated by a Synthetic Selection Strategy. *ACS Synth. Biol.* (2018) doi:10.1021/acssynbio.8b00215.

199. Cui, Z., Jiang, X., Zheng, H., Qi, Q. & Hou, J. Homology-independent genome integration enables rapid library construction for enzyme expression and pathway optimization in Yarrowia lipolytica. *Biotechnol. Bioeng.* (2019) doi:10.1002/bit.26863.

200. Wagner, J. M. & Alper, H. S. Synthetic biology and molecular genetics in non-conventional yeasts: Current tools and future advances. *Fungal Genet. Biol.* (2016)

doi:10.1016/j.fgb.2015.12.001.

201. Schwartz, C. M., Hussain, M. S., Blenner, M. & Wheeldon, I. Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR-Cas9-Mediated Genome Editing in Yarrowia lipolytica. *ACS Synth. Biol.* (2016) doi:10.1021/acssynbio.5b00162.

202. Schwartz, C., Shabbir-Hussain, M., Frogue, K., Blenner, M. & Wheeldon, I. Standardized Markerless Gene Integration for Pathway Engineering in Yarrowia lipolytica. *ACS Synth. Biol.* **6**, 402–409 (2017).

203. Schwartz, C. & Wheeldon, I. CRISPR-Cas9-Mediated genome editing and transcriptional control in yarrowia lipolytica. in *Methods in Molecular Biology* vol. 1772 327–345 (2018).

204. Rutter, C. D. & Rao, C. V. Production of 1-decanol by metabolically engineered Yarrowia lipolytica. *Metab. Eng.* (2016) doi:10.1016/j.ymben.2016.07.011.

205. Lomakin, I. B., Xiong, Y. & Steitz, T. A. The Crystal Structure of Yeast Fatty Acid Synthase, a Cellular Machine with Eight Active Sites Working Together. *Cell* (2007) doi:10.1016/j.cell.2007.03.013.

206. Leibundgut, M., Maier, T., Jenni, S. & Ban, N. The multienzyme architecture of eukaryotic fatty acid synthases. *Current Opinion in Structural Biology* (2008) doi:10.1016/j.sbi.2008.09.008.